Intelligent Holographic Microscopy: identifying blood cells without labelling

TaoXu

U5527268

COMP 8755

Supervised by Dr W M (Steve) Lee

Examiner Miaomiao Liu

A thesis submitted for the degree of

Master of computing of

The Australian National University

June 2020
This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where the reference has been made in the text.

Tao Xu

May 2020
Table of Contents

1 Abstract .................................................................................................................. 4

2 Introduction ........................................................................................................... 4
  2.1 Previous Work .................................................................................................. 5
  2.2 Digital Holographic Microscope and Deep learning ...................................... 6

3 Methods ................................................................................................................. 7
  3.1 Generative adversarial networks ...................................................................... 7
  3.2 Mask Region-based Convolutional Neural Network ....................................... 7
  3.3 Phase reconstruction ....................................................................................... 8
    3.3.1 Data acquisition ........................................................................................ 9
    3.3.2 Data pre-processing .................................................................................. 10
    3.3.3 Pix2Pix GAN for phase reconstruction ...................................................... 12
  3.4 Object detection ............................................................................................... 15
    3.4.1 Data pre-processing .................................................................................. 16
    3.4.2 Mask R-CNN neural network for object detection .................................... 16

4 Result and discussion ............................................................................................ 18
  4.1 Phase reconstruction Results .......................................................................... 18
  4.2 Object detection Results .................................................................................. 19

5 Conclusion .............................................................................................................. 20

6 Reference ................................................................................................................ 20

Appendix A: .................................................................................................................. 22
Appendix B: .................................................................................................................... 25
1 Abstract
The neural network has been widely used in the image processing area such as object segmentation, object detection, image reconstruction and so on. Recently, more and more neural networks were used in biological imaging such as cell classification, cell tracking and cell edge detection in phase-contrast microscopy. However, some of phase-contrast microscopy use thresholding method to do phase reconstruction and cell segmentation. This thesis focuses on red blood cell (RBC) phase image reconstruction and RBC detecting through Generative adversarial network (GAN) and Mask Region-Based Convolutional Networks (Mask R-CNN). GAN achieved phase image reconstruction and Mask R-CNN achieved RBC segmentation.

2 Introduction
Red blood cells (RBC) are indispensible to the human body, which delivers oxygen to other organs through haemoglobin. There are many kinds of microscopes are used to look at cells and study their physiological mechanisms such as bright field microscope, dark field microscope, fluorescence microscope, phase contrast microscope, laser scanning confocal microscope, polarizing microscope. In this thesis, the digital holographic microscope was used to collect the RBC images and a MATLAB code reconstruction software was used to generate the reconstructed phase image. The Figure 1 shows the setup that I used in this project. For the current setup, it has achieved phase reconstruction and cell segmentation through thresholding method [1].However, it needs prior knowledge, filtering operation, phase aberration and unwrapping processes, which includes many complex steps. Holographic imaging is an effective technique to record diffracted wavefront that includes the amplitude and phase information[2]. The amplitude and phase information can be numerically reconstructed by using a computer, which represented the three-dimensional(3D) image. The Fresnel-Kirchhoff integral is the common theory to do numerically digital holographic (DH) reconstruction [3]. The principle of the DH is using interferometry and Fourier optical transform to measure phase shift. The interferometry pattern can encode the phase information and interference fringes illustrate the phase changes as disturbances. The double-slit interference experiment that is demonstrated by Sir Thomas Young lead to this principle [4]. Numerically reconstruction method can decode the phase information from the interferometry pattern. Moreover, there are many kinds of digital holography configurations such as off-axis Fresnel Holography, Fourier Holography, image plane holography, in-line holography, Gabor holography and phase-shifting digital holography [5]. Therefore, different designs of Digital Holographic Microscopy (DHM) were generated by using these digital holography configurations. With the classical imaging cell culture plates, the DHM can observe transparent cells. The intracellular refractive index and cell thickness have a strong link with the contracted phase image. In the biological research area, it is important to observe live biological sample morphology. Therefore, the microscope is the main tool to observe living cells. Live cells are a very complex system. It contains many things and related to cellular mechanics. As the complex cellular dynamics, it is necessary to develop a three-dimensional imaging system as a tool to help biological studies. Consequently, there are lots of work have been done for the three-dimensional imaging microscopy. Laser scanning technology enhanced the development of real-time three-dimensional imaging. A series of fluorescence microscopy has achieved development from immaturity to maturity, such as confocal microscopy[6], light-sheet microscopy[6], multiphoton microscopy [7]. However, there are some disadvantages to fluorescence microscopy. First of all, it needs to selectively stain the sample and sample preparation needs lots of manpower and time. Moreover, the fluorescence molecules may affect the cell original morphology. Secondly, when do the investigation, the laser will excite the fluorescence directly
and it is not able for long time observation. These drawbacks can be tackled by Digital Holographic Microscopy which is phase-based light microscopy. DHM is a label-free imaging technique, which is a powerful tool to do label-free live cell investigation. It can retrieve the phase delay when the light passes through a sample and then generate a height information of the sample. Finally, a three-dimensional information was calculated by the digital reconstruction algorithm. Moreover, the most of microscopes has applied machine learning technique for image processing. Especially, DHM uses machine learning technology to do phase reconstruction and object detecting. Object detection technique development is based on image classification. The input is a training set composed of N images, with a total of K classes, and each image is marked as one of them. Then, use the training set to train a classifier to learn the external characteristics of each category. Finally, we predict the class label of a new set of images and evaluate the performance of the classifier. We compare the category label predicted by the classifier with its real category label. The current popular image classification architecture is a convolutional neural network (CNN), which feeds images into the network and then the network classifies the image data. The convolutional neural network starts with the input "scanner", which does not parse all the training data at once. For example, you don't need a layer with 10,000 nodes to input an image with a size of 100 by 100. Instead, you only need to create a scan input layer of size 10 by 10, scanning the first 10 by 10 pixels of the image. Then the scanner moves one pixel to the right and scans the next 10 by 10 pixels, which is the sliding window. The input data is fed into the convolutional layer instead of the normal layer. Each node only needs to deal with its nearest neighbour, and the convolutional layer tends to shrink with the deepening of the scan. In addition to the convolution layer, there is usually a pooling layer. Pooling is a method of filtering details. A common pooling technique is maximum pooling, which uses a 2 by 2 matrix to pass the pixels with the most specific attributes. Currently, most image classification techniques are trained on the ImageNet dataset, which contains about 1.2 million high-resolution training images.

2.1 Previous Work

For the currently Set up (Figure 3), it has achieved phase reconstruction and cell segmentation through threshold iteration and watershed methods. For the phase reconstruction, the Figure 1 [8] shows the process of phase reconstruction. Using fast Fourier Transform to get the frequency domain from hologram (a) to (b). Then using threshold iteration to generate three frequency domain areas. After that using iFFT to get the phase information and phase unwrapping and phase aberration correcting were applied and get the phase image.

![Figure 1 Phase Reconstruction process](image)
The Figure 2 [8] shows the process of cell segmentation. It loads the phase image first and apply gaussian filter on the phase image. Next step extract bounding pixels and fill holes and remove small objects. Then to get the location of individual cells. After that inverse intensity of the image. Using watershed transform to detect the cells. Finally, labelling cell regions and detect the cells.

![Figure 2 Cell Segmentation Process](image)

As mentioned above, there are many complex steps to achieve these two functions. Hence, in my thesis, I demonstrated using deep learning neural networks to achieve these two functions. The first neural network will take hologram image as input and get the phase image as output. The second neural network take phase image as input and get the cell segmentation phase image as output.

### 2.2 Digital Holographic Microscope and Deep learning

DHM has been achieved by different functions by using deep learning. The deep learning helps DHM get the aberration-free quantitative image as the traditional DHM always tackle the phase aberration compensation issues through manually detecting the background for quantitative measurement [9]. Moreover, DHM reconstruction is also achieved through the end to end deep learning method because the original DHM reconstruction needs to know object distance, the incident angle between the two beams, and the source wavelength and also need to filter the zero-order and twin images, which consumes more time in off-axis configuration [10]. Furthermore, the autofocus function was achieved by deep learning, which is very useful for multiple sectional objects. Autofocusing uses entropy or variance to calculate the sharpness of reconstructed in focus and out of focus images, which is computationally and time-consuming. Therefore, deep learning converts the autofocus problem to classification problem [11]. In addition, the deep neural network achieved the digital staining through the generative adversarial network with paired images [12]. Stain-free is a big benefit for biology research as staining cell is time-consuming and waste the labour. Some of the research group has used the MaskR-CNN achieved the stain-free and single-cell segmentation [13]. Besides, convolution neural networks (CNN) has been used for RBC classification based on the RBC shape features [14]. In this thesis, the RBC phase retrieval and RBC image segmentation and stain were demonstrated together through deep learning.
3 Methods

This part described the way that was used for phase reconstruction and object detection. Before applying the neuro network to do that, the data preparation and pre-processing were demonstrated first. The phase reconstruction data was generated by using a Matlab application which is written by our lab PhD student. All the raw data of the red blood cell and platelets are video file recorded by a CCD. Every frame of the video is a hologram image that contains intensity and phase information. The Matlab application can extract every frame from the video and reconstruct the phase image by using numerically reconstruction method. After that, the laboratory digital holographic expert applied a data wrangling skill on the generated data to pick out high-quality phase image data that was used for training data set of phase reconstruction. The phase reconstruction is achieved by Pix2Pix Generative adversarial network (pix2pix GAN) [15]. The pix2pix GAN can generate phase image that was used for object detection through Mask Region-based Convolutional Neural Network (Mask R-CNN)[16]. This object detection networks detect red blood cell and platelets from the phase images based on the object detection training dataset. The training phase images are labelled by Labelling tool that is written by Python with Qt user interface. When labelling finished, it stores the files as XML that can be used in ImageNet as training data set. The training dataset was used to train the neural network.

3.1 Generative adversarial networks

Pix2pix GAN is one type of generative adversarial networks (GAN). GAN is an unsupervised machine learning method, which allows two neural networks to game each other and optimise the parameters[17]. Generally, the GAN has two networks and they are generator and discriminator. The generator produces a result from latent space and aims to generate a fake result which closes to the real target. The discriminator would evaluate the result of the generator based on the training dataset. Hence, the parameters were optimised after serval iterations and the discriminator cannot judge the if the result of the generator is real or fake. The GAN has been used for many applications. GAN can be used to create art photos, fashion models without a professional photographer or makeup artist. Also, it facilitates the research of science such as astronomy image features recover[18] and biological stain imaging[19]. The GAN only need the backpropagation to update the parameters and it does not need the sample to update the networks. However, the GAN generator produces result extremely random without any pre-built model and generating good result only when generator and discriminator are balanced. Consequently, conditional generative adversarial networks (cGAN) was developed, which does not involve stochasticity in a generator [20]. The pix2pix GAN is a conditional generative adversarial network. The pix2pix GAN could generate the phase image from hologram image.

3.2 Mask Region-based Convolutional Neural Network

For the object detection neural network, the Mask Region-based Convolutional Neural Network (Mask R-CNN) is a state of the art approach [16]. Since 2017, the single-task network structure has gradually ceased to lead the object detection and is replaced by an integrated, complex, multi-tasking network model. The Mask R-CNN is a typical representative. The Mask R-CNN marked object by using the bounding box and classified every object in a specific class, and it achieved pixel-level segmentation[16]. The Mask R-CNN inherited from Faster R-CNN[21], and Faster R-CNN inherited from Fast R-CNN[22], and Fast R-CNN is inherited from R-CNN[23]. The convolutional neural network has traditionally been applied in spatial
problem domains[24]. It is a class of deep neural networks and was popularly used in images domain, temporal domains and sequential data. In this thesis, the network works with image data. For the R-CNN, it extracts the number of regions first. Then the CNN compute the features of each region. These features will be classified through SVM and generate the object class. For the first step extracting region proposals, it using selective search for object recognition[25]. The core algorithm of selective search is SVM. The features computing is achieved by using AlexNet which is trained by image net. However, the R-CNN selective search is time-consuming for each image, which is around 2s per image. Serial CNN forward propagation is also time-consuming as every region of interest(RoI) features use Alexnet to extract and each one costs 47s [23]. Fast R-CNN changed the serial CNN and extract features directly from the whole image. Except for selective search and other parts can train together. Therefore, Faster R-CNN was developed. The Faster R-CNN removed the selective search method and use Region Proposal Network (RPN) to generate detection regions. The Faster R-CNN use the shared convolutional layer to extract the features for the whole image and then features map was sent to RPN. Then the RPN generate the detect region and perform the first correction of RoI bounding box. After that, the Fast R-CNN architecture is applied. Based on the RPN output, the RoI pooling layer select the features corresponding to each RoI on the feature map. Finally, the fully connection layer was used to classify the image and perform the second correction of the target image. The Faster R-CNN is truly implementing end-to-end training. However, the Faster R-CNN RoI pooling is using rounding method, which is bad for pixel-by-pixel prediction result because the features obtained for each RoI is not aligned with the RoI. Hence, the Mask R-CNN improved the RoI pooling method and proposed RoI align method. The role of RoI Align eliminated the rounding operation of RoI Pooling and make the features of each RoI to align the RoI area on the original image.

3.3 Phase reconstruction

Frits Zernike demonstrated the phase contrast microscopy in 1934[26]. This microscopy was widely applied because it provides high contrast of transparent thin biology sample. Therefore, there are some techniques was developed such as Nomarski microscopy (NIC) [27] and Hoffman modulation contrast microscopy (HMC) [28]. The NIC generates similar image of phase contrast microscopy but without bright diffraction halo. Comparing with NIC, HMC increased the contrast through optical component in the light path. These two techniques were focus on qualitative, which do not provide specific phase changes or path difference. As mentioned above, Gabor introduced holography which records amplitude and phase in an image. It means the entire light filed was recorded. After that, the development of lasers promoted the phase microscopy researches. Quantitative phase microscopy (QPM) was constructed, which is general name for a group of microscopies. These microscopes can measure the phase delay by using formula (1). \( \Delta n \) is the index difference between sample and medium. \( \lambda \) is the wavelength of the laser. \( h \) is the thickness of the sample. There two main methods to achieve QPM, off-axis and common-path methods. For this thesis uses the off-axis set up to generate the data, which is shown in Figure 3 [8]. Figure 1 shows the experiment set up inverted DHM microscope. It uses a continuous wave laser (\( \lambda = 632.8\)nm). The output of the laser was focused into a single-mode optical fibre by using a microscope objective (MO). After the MO, there is an optical fibre splitter that is used to separate the laser into two beams (object beam and reference beam). The object beam was focused by a lens and the focus point is on the back focal plane of the second microscope objective (MO2). The next microscope objective (MO3) will collect MO2 output object information. The reference beam was
expanded by two lens. Finally, the object beam and reference beam were combined through a non-polarizing beam splitter. The combined beam was passed to the charge couple device (CCD) camera. The CCD could record the phase information by using an interferometry configuration as the DHM principle. The interference fringes could display the phase information as disturbances. \( R_0 \) and \( O_0 \) shows the reference wave and object wave in formula (2) and (3). The \( \omega_1 \) menas the angular frequency of reference wave and object wave. \( \varphi \) in (3) is the phase change produced by the RBC sample. Based on the wave formulas, the hologram pattern formula can be written as formula (4).

\[
\Delta \varphi = \frac{2\pi \Delta n}{\lambda h} \quad (1)
\]

\[
R(r,t) = R_0 \exp[j(-\omega_1 t + (\vec{k} \cdot \vec{r}))] \quad (2)
\]

\[
O(r,t) = O_0 \exp[j(-\omega_1 t + (\vec{k} \cdot \vec{r}) + \varphi)] \quad (3)
\]

\[
I_n = O_0^2 + R_0^2 + |O_0 R_0| e^{j\varphi} + |R_0 O_0| e^{-j\varphi} \quad (4)
\]

**3.3.1 Data acquisition**

The RBC sample was prepared by the lab PhD students. They flow the RBC solutions through a Polydimethylsiloxane (PDMS) microchannel with a pump. The PDMS microchannel was fabricated by soft lithography techniques and the fabrication process as shown in Figure 4. Figure 4 a) shows the process of the soft lithography. The process uses negative SU8 photoresist to make the channel on the wafer. First of all, the wafer should be cleaned by using deionized water (DI) and Isopropyl Alcohol (IPA) solution, which aim to remove dust and fingerprint. After that, the wafer will put on the baking machine to remove the moisture at 200°C. The next step is coating the photoresist on the wafer through the coating machine. The coating machine identified the thickness of the photoresist on the wafer, which is the height of the microchannel. After coating the photoresist, the wafer put on the baking machine again for solid the photoresist on the wafer. It takes 1 hour for baking. The photolithography mask applied on the baked wafer through the UV machine. After UV light passing the mask, the photoresist reacts with the UV and form cross-link. Following this step, the wafer needs to bake again and to make the photoresist could be solvable in developing solution. It takes 2 hours to
do the hard bake. After 2 hours, making the wafer cooling down and then put it into the developer solution. Shaking the developing solution and remove the solvable photoresist and left the pattern on the wafer which is shown in Figure 4 c). Then pouring the PDMS on the wafer and put it into the oven and make it forming a channel on the PDMS. Taking out the PDMS and cut it. The independent PDMS channel was generated and attach on the glass slides, which is shown in Figure 4 d). Figure 4 b) shows the difference between negative and positive photoresist. The negative photoresist will become solid after the UV light applied on the photoresist. On the contrast, the positive photoresist will become solvable after UV light applied on the photoresist. Figure 4 d), shows the microchannel that has two holes on the channel, which is used for pump RBC solution. Put microchannel chip under the objective. Next step to adjust the objective and focus on the channel base as the RBC flow along the base surface. Then start the pump to flow the RBC solution. The CCD has the software to observe the channel flow and move the stage by suing XY stages to choose the region of the channel and wait for RBC to enter the viewing region. Once got the RBC image and click the record button to record the RBC flowing process. There is also an open-source Matlab tool to collect the images and record the video. The Matlab tool could record the interference image and store them as AVI file. The software was shown in Figure 3.

**Figure 4 Soft-lithography process**

### 3.3.2 Data pre-processing

The AVI file of the RBC video recording was generated. The next step is preparing the training data for training the neural network. The training data includes source image and target image. The source image is every frame of the RBC video, which is the hologram image. The target image needs to be reconstructed from hologram and to get the phase image. There is a self-designed user-friendly Matlab code software to reconstruct the hologram to phase image. This software was designed by PhD student. The user interface was shown in Figure 5 a). The
program has recording function that I have mentioned in data acquisition section. According
to the user interface showing, there are two main parts. On the left is imaging function which
connect to the CCD. It can set parameters for recording such as camera mode, exposure time,
file name, live view frame rate and recording frame rate. The processing part is the main
function of this software. It can read the video file and extract each hologram frame and
reconstruct it. Firstly, press the select file button and select the video file. The program will
read the video file and extract frames from it. The preview button is used to preview the
recorded video. Secondly, users can select ROI by using the set ROI button. Once user pressed
the button, a new window pops up (Figure 5 b) and a crosshair will appear. The user can move
the crosshair on the image and select ROI through cropping the image. Also, user need to set
parameters for processing the image. The pixel size, wavelength and refractive index setting
were from the hardware. Pixel size depend on CCD, wavelength depend on lasers and the
refractive index depend on the solution that we use. Thirdly, after select a ROI, the run manual
and run auto buttons were activated. Run auto will select the first order automatically and run
manual select the first order by users. If user press selects manual, another window pops up
(Figure 5 c) and then select the first order. After selecting the first order, the program starts to
process the hologram. The Figure 6 shows the result of the software. Based on the result, we
can see that the result includes hologram, intensity, intensity curve removed, phase, phase
unwrapped, phase unwrapped curve removed, spatial frequency cropping, thickness and
thickness plot. The hologram image and phase unwrapped curve removed image were training
data that we used in training network. However, the phase unwrapped curve removed image
need to convert to grayscale for easy training. The phase unwrapped curer removed matrix has
the value range from 0-10 with 4 decimal places so that the matrix needs to multiply 10000 for
accurate grayscale image and then the grayscale matrix saves as a jpg grayscale image. In
addition, the video ROI only focus on one region. Therefore, the limited training data was a
problem for training the neural network. To tackle this problem, the hologram image needs to
crop as different small hologram images and then reconstruct each hologram to phase image,
which is used for training the network. A self-written Matlab code was applied on these
hologram images through a for loop. The code read the hologram image using imread() function and change the RGB to grayscale using rgb2gray() function. The fftshift() function was used to translate to centre through the 2D Fourier Transform. The FFT image using gaussian fit and get the global threshold level. Then remove the noise region and only left order region. Generating the filter mask based on the selected order region and apply mask. Then to transfer the FFT region on a black image and the mask will fit on this image for phase unwrapping. The final step is intensity and phase reconstruction through inverse Fourier Transform. The intensity can be reconstructed by using inverse FFT but the phase reconstruction need to transfer the inverse FFT to angle and multiply the invert factor. To get the phase unwrapped and no curve image, the reconstructed phase needs to minus the curve phase. Finally, there are around 23 paired training data for the neural network.
3.3.3 Pix2Pix GAN for phase reconstruction

The training data was prepared after data pre-processing. Next is to apply the training data on the pix2pix GAN network. As mentioned before, the GAN neural network has two main parts: generator (G) and discriminator (D). G is a role for generating images. After input a random code Z, it will output a automatically generated fake image G(Z) through the neural network. The D is used for checking the output of G and check if the image is real or fake, if it is fake the output of D is 0 and otherwise is 1, which shown in Figure 7. In the training process, the two networks play games with each other. Both networks become more and more capable. The image generated by G becomes more and more authentic, and D becomes more and more able to judge the authenticity of the images. At this point, we can get rid of D and use G as an image generator. The formula (5) shows that on the premise of maximizing the ability of D, and minimize the ability of D to judge G, which is a minimum and maximum problem. In order to enhance the capabilities of D, we consider the case of input real image and fake image.
respectively. Based on the formula (5), the first item $G(Z)$ is for processing the fake image then the score $D(G(Z))$ need to do the best to reduce. The second item $(1-D(x))$ deals with real image $X$, where the score is higher. However, the traditional GAN does not have user control ability as it always uses random noise to generate image. Moreover, the traditional GAN image has low resolution and low quality. For solving these problems, the pix2pix GAN was developed. The pix2pix GAN only edit a part of the traditional GAN. The $G$ will not edit a lot. The input of $D$ was changed. The traditional GAN $D$ only take $G$ output as input and the pix2pix GAN take the target image, $G$ output and source image as input together so the $D$ can determine if the image is real or fake by comparing the target image and the fake image. Figure 9 shows the following chart of the pix2pix GAN. Hence, the pix2pix GAN network needs paired dataset that includes source image and the target image. Figure 8 a) shows an example for the dataset and Figure 8 b) shows the source image (hologram image) on the left and reconstructed phase image on the right. In this thesis, the hologram image is source image and the phase image is the target image.

$$\min(G) \max(D) E[\log D(G(z)) + \log(1 - D(x))]$$  \hspace{1cm} (5)$$

**Figure 7 GAN Network**

In this thesis, the $G$ is more complex than the $D$. The $D$ implements using the 70*70 PatchGAN model [15]. This model uses two images as input that are concatenated together and predicts the predictions’ patch. The binary cross-entropy was used to optimize the method. The weighting of this model updates should slow down relative to the generator model during training. The $G$ is an encoder-decoder U-net architecture. It takes the source image to generate the target image. U-net architecture was shown in Figure 8. U-net is an image segmentation technique. The U-net is based on fully convolutional neural network (FCN). FCN uses upsampling and deconvolution to the original image size and then do the pixel-level classification. Based on Figure 8, the U-net was divided into two parts. The left part is used to extract the features and the right part was used to up sampling. In the up-sampling part process, every up sampling the number of channels corresponding to the feature extraction part was combined at the same scale but it needs to crop before combining.
According to Figure 10, the input is a 572*572 image which is on the left part. Also, the left part is called contracting path and it includes 4 blocks. There are 3 blocks use convolutional and 1 block uses max pooling. Every downsampling, the feature map will be increased by double so finally, it gets a 32*32 feature map. On the right part, it is called expansive path. It also includes 4 blocks. Before every block starting, the feature map will multiply by 2 through deconvolutional method. Then it will combine with the feature map on the left as mentioned before. The output will be a 388*388 feature map. The formula (6) shows how does the U-net detect the edge by using the loss function. The $P_l(x)(X)$ is the softmax loss function and $l: \Omega \rightarrow (1, \ldots, K)$ is pixel tag value, $\omega: \Omega \subseteq \mathbb{R}$ is the pixel weight, which used for giving higher weight to the edge pixel.

$$E = \sum_{x \in \Omega} \omega(x) \log (P_l(x)(X))$$  \hspace{1cm} (6)

There are different types U-net for image segmentation such as 3D-Unet, ternausNet, Res U-net and multiRes U-net and so on. They all use convolution and deconvolution technique to achieve image segmentation.
3.4 Object detection

Object detection in an image usually involves outputting bounding boxes and labels for the individual project. This is different from the classification/localizing task, which is to classify and locate many objects, not just individual subject objects. In object detection, you only have two object categories, object bounding boxes and non-object bounding boxes. For example, for car detection, you must use a bounding box to detect all the cars in a given image. If we use a sliding window technique like image classification and image positioning, we need to apply the convolutional neural network to many different objects on the image. Since the convolutional neural network will recognize every object in the image as an object or background, we need to use the convolutional neural network at a large number of locations and scales, but this requires a large amount of computation. Therefore, region convolutional neural network (R-CNN) was developed. In R-CNN, a selective search algorithm is first used to scan the input image for possible objects, generating about 2,000 area suggestions. Then, a convolutional divine network is run over these region suggestions. Finally, the output of each convolutional neural network is fed to a support vector machine (SVM), which uses linear regression to tighten the object's bounding box. However, training is slow, requires a lot of disk space, and detecting is slow. Consequently, Fast R-CNN was developed. For Fast R-CNN, feature extraction is carried out before the proposed region, so the convolutional neural network can only be run once on the whole image. Instead of creating a new model, use a softmax layer instead of a support vector machine to extend the neural network used for prediction. Figure 11 a) shows the R-CNN and Figure 11 b) shows the Fast R-CNN.

Figure 11 R-CNN and Fast R-CNN
3.4.1 Data pre-processing
The phase data was generated by the pix2pix GAN networks. For object detection, it needs the original phase image and label phase image. To label the image, the python labelme tool was used. The tool user interface was shown in Figure 12 a) and one labelled image example was shown in Figure 12 b). According to the Labelme user interface, there are open, open dir, next image, previous image, save, create a polygon and edit polygon. Firstly, click the open dir and select the folder where the annotation files are located, and start the annotation. For example, if the object you want to mark is human and dog, in the process of labelling, the multiple persons or dogs each of them has an individual label. The naming rules are person1, person2…Dog1, Dog2… Because Labelme generates a label.png file, which has only one channel, the same label mask will be given a label bit when you label, and the mask requires different instances to be placed in different layers. The input required for the final training is a w*h*n array, where n is the number of instances in the picture, w is the width and h is the height of the image.

![Figure 12 Labelme user interface](image)

3.4.2 Mask R-CNN neural network for object detection
Mask R-CNN is a very flexible framework, which can add different branches to complete different tasks, including object classification, object detection, semantic segmentation, instance segmentation, body posture recognition and other tasks. Mask R-CNN has high speed, high accuracy, simple and intuitive and easy to use properties. For the high speed and high accuracy, Faster R-CNN and classical semantic segmentation algorithm FCN combined to generate Mask R-CNN. Fasters-R-CNN can achieve the function of object detection quickly and accurately. FCN can accurately complete the semantic segmentation function. Mask R-CNN is more complex than Faster-R-CNN, but it can still reach the speed of 5fps eventually, which is similar to the speed of the original Faster R-CNN. As the pixel deviation problem in ROI Pooling was found, the corresponding ROIAlign method was proposed, and the accurate pixel mask of FCN was added, which can achieve high accuracy. The idea of the whole Mask r-cnn algorithm is very simple. FCN is added based on the original Faster R-CNN algorithm to generate the corresponding mask branch. The mask R-CNN can represent as RPN + ROIAlign + Fast-R-CNN + FCN.

The Mask R-CNN algorithm is simple for implement. First, enter the image you want to process, and then do the corresponding pre-processing operation, or after the pre-processing of the image. Then, it is input into a pre-trained neural network to obtain the corresponding feature map. After that, a predetermined ROI is set for each point in the feature map to obtain multiple
candidate ROIs. These candidate ROIs were sent to the RPN network for binary classification (foreground or background) and bounding box (BB) regression to filter out some candidate ROIs. Performing ROIAlign operation on the remaining ROIs. Finally, these ROIs were categorized (N categories), BB regression, and MASK generation.

Mask R-CNN can be decomposed into the following three modules: Faster-R-CNN, ROIAlign and FCN. As above mentioned, the Figure 11 b) shows the Fast R-CNN process. Firstly, the input image is cropped, and the cropped image is fed into the pre-trained classification network to obtain the corresponding feature map of the image. Then on the characteristic image of each anchor point take some candidate ROI and map it to the corresponding proportion in the original image, which is for feature extraction of network commonly convolutional and pool, but only the size of the pool will change the characteristic image, so the final figure related to the size and the number of the pool. Then enter the ROI of the candidate to the RPN network, RPN network to classify the ROI at the same time carries on the preliminary regression, and then do maximum inhibition. After that, carrying out ROI Pooling operation for these ROIs of different sizes, and output feature_map of fixed size. Finally, fed it into a simple detection network, and then classified by the convolution of 1x1. Meanwhile, BB regression is carried out, so output a BB set. FCN algorithm is a classical semantic segmentation algorithm, which can accurately segment the object in the image. Its overall architecture is shown in Figure 13 a). It is an end-to-end network. The main modulo speed includes convolution and deconvolution. Then, the deconvolution operation and interpolation operation is carried out first and the size feature map is constantly decreased. Finally, the value of each pixel is classified. Thus the accurate segmentation of the input image is realized. Figure 13 b) shows the ROIAlign. To obtain the fixed-size feature map, the ROIAlign technology does not use quantization operation, so there is no quantization error. For example, 665/32 = 20.78, just use 20.78, will not use 20 to replace it. This is the original intention of ROIAlign. Hence, dealing with these floating-point numbers, the bilinear interpolation algorithm was implemented. Bilinear interpolation is a relatively good image scaling algorithm, it fully uses of the virtual point in the original image (such as 20.5 the floating-point number, the pixel position is an integer value, no floating-point value) around four real pixels to jointly determine the target image in pixel value, namely can be 20.56 corresponding pixel values of the virtual position estimation. As shown in Figure 13 c), the blue dotted box was generated after the convolution of the feature of the map. The black solid line boxes represent ROI feature. For the 2x2 output, it will use bilinear interpolation to estimate the blue dot place corresponding pixel values and then get the appropriate output.

The ROI loss calculation changed due to the mask branch was implemented. Every ROI loss has shown as formula (7). L_cls and L_box have the same definition in Faster R-CNN. For each ROI, the mask branch has the output with the K(m*m)dimension, which encodes the number of K masks, each with K categories. It used per-pixel sigmoid and defined L_mask as the average binary cross-entropy loss. L_mask is only defined on the Kth mask. The L_mask definition allows the network to generate a mask for each class without competing with other classes. It relied on the predicted category label by the classification branch to select the mask. This separates the categories from the mask generation. This is different from FCN semantic segmentation, FCN usually uses a per-pixel sigmoid and a multinomial cross-entropy loss, there is a competition between masks in this situation. However, per-pixel sigmoid and a binary loss were used, there was no competition between different masks, which can increaser the
segmentation performance. Specifically, FCN was used to predict an m×m size mask from each ROI, which enabled each layer in the mask branch to explicitly maintain an m×m spatial layout without folding it into vector representations that lacked spatial dimensions. Different from the previous method of using the FC layer as mask prediction, the mask representation needs fewer parameters and is more accurate. These pixel-to-pixel behaviours need ROI features, and ROI features are usually a small feature map, which has been processed. To maintain a clear spatial correspondence of single-pixel consistently, the ROIAign operation comes out.

\[ L = L_{cls} + L_{box} + L_{mask} \quad (7) \]

Figure 13 FNC and ROIAign

4 Result and discussion

4.1 Phase reconstruction Results

After data pre-processing, the paired data (hologram and phase image) for GAN neural network input was generated. The image size is 256x256. These data were used for training the GAN and finally, the generator was used to generate the phase image from the hologram. The output is a grayscale phase image data. Figure 14 shows the result of the generator at 23 epochs, which includes source, generated and expected images. According to the result, the phase images look very realistic and close to the expected image after 1000 epoch. Also, the Figure 14 a) shows the losses of the GAN, which includes loss of real example discriminator(d1), and loss of fake example of discriminator (d2) and loss of generator(g). The d1 loss is 0.029 and the d2 loss is 0.020 and the g loss is 46.306. The GAN aims to get a generator with a low loss, which means the generator can produce a high-quality image having the same value of each pixel of the
expected image. Hence, a higher number of training epochs was applied. Figure 14 c) shows the result of after 1000 epochs. The d1 is 0.283 and d2 is 0.164 and the g is 3.723. The loss value of g dropped a lot comparing with 46.306. Furthermore, based on Figure 14 d), the image quality also increased a lot, which cannot tell by eyes if the image is fake or not.

![Figure 14 Result of the generator](image)

When the loss of the generator dropped to 3.178 at 1035 epochs, the training stopped. And the image result of training data performs very well at this point. The Figure 15 shows the trend of the generator loss.

![Figure 15 Generator loss plot](image)

4.2 Object detection Results

For object detection, the Mask R-CNN neural networks were trained by the red blood cell phase images. The trained Mask R-CNN can achieve three main points in the result. Firstly, it can detect the object and draw a bounding box on the resulting map. Secondly, object classification for each object, the corresponding class should be found to distinguish whether it belongs to a category. Finally, it can achieve pixel-level segmentation. In each object, it is necessary to distinguish at the pixel level, what is the foreground and what is the background. The training data red blood cell phase image has two classes, red blood cell and background. The R-CNN will use training data to train the neural network and predict the RBC in a new image. The detected RBC was marked as red colour and with an outline of the RBC. Also, there is another
result which shows the probability of the RBC with a tag bounding box. Figure 16 a) shows the input image and Figure 16 b) and c) show the two result. The (b) shows the stained RBC with outline and (c) shows the stained RBC, bounding box, label and percentage of the RBC.

![Figure 16 Mask R-CNN Results](image)

The watershed and Cell Tracker methods were implemented to compare the difference between Mask R-CNN. The Figure 17 shows the Mask R-CNN, Cell tracker and Watershed difference in detail.

![Figure 17 Result Comparation](image)

According to the Figure 17, the Mask R-CNN can detect the overlapping RBC clearly. The watershed method detects the overlapping RBC as one cell, which same as the cell tracker results. Moreover, the watershed missed one RBC when it detects the cells. However, we can see the cell tracker can detect more RBC. The main reason for the Mask R-CNN can not detect more RBC is limited training data. The COVID-19 event closed the school, I cannot go campus to collect more training data to train my neural network.

5 Conclusion

Red blood cell is an important part of the human body, which maintained the everyday functions of the body by transporting oxygen. RBC dynamic shapes changing is related to aetiopathogenesis and detecting the shape of the RBC is a crucial step to diagnosing. Furthermore, RBC detecting also provided an easy way to collect image data for researching.
In this thesis, a digital hologram microscope was demonstrated, which can collect 3-D RBC information based on phase shifting. Therefore, a Generative adversarial network was implemented to do the phase image reconstruction, which improved the traditional reconstruction method. For the RBC detecting, the Mask R-CNN was introduced, which detected RBC with the bounding box and, it also shows the outline and probability of RBC. The Mask R-CNN provided a reliable RBC detecting result, which shows the clear location of the phase image RBC and could help to reconstruct the phase image accurately. These two networks were renovated on the existing networks and the vital step of implementing is data pre-processing.

In the future, more training dataset will be acquired, and better training machine will be used, which can improve the performance of the networks. Moreover, the data pre-processing plays an important role in this thesis. Most of work focuses on data pre-processing so that the data pre-processing technique need to be improved, which can help to train a high-performance neural network. After the neural networks having an ideal performance, to track the dynamics of RBC inflow will be applied, which can observe the real-time changing of the RBC and detect the living RBC deformation in the flow. Different shapes RBC represent the different situations of the vessel and also it related to the vessel wall pressure, which could help to diagnose some kinds of vascular disease based on the vessel shape-changing and vessel wall pressure changing.

6 Reference

**Appendix A:**

Phase reconstruction:

1. For the phase reconstruction it includes phase_reconstruction.ipynb, which could be open through Jupyter notebook.
2. There is a folder that includes three matlab code files. The first one reconstructed_new was used to reconstruct the phase image. And the rename code was used to rename the file to respond training data name.
3. The tif to jpeg was used to convert the tif image to jpeg as the original image type is tif.
4. Video extract can extract frames from video to get the hologram image.

Run the program: to change the path of Source folder and target2 folder.
The user can generate the own training data using the reconstructed_new.m file.

Cell segmentation:

It includes model, utils, visualize, paralle_model, config and RBC python file.

1. Open the command window.

2. Using cd command direct to RBC python file folder
3. Using “python3 RBC.py train --dataset=/path/to/dataset --model=coco” to train the model. As the limited training data, so transfer learning was used. It uses image net coco weight to train the network. The trained weight will store in logs folder.

4. Using the trained weight to detect the cells. “python3 RBC.py detect --weights=/path/to/mask_rcnn/mask_rcnn_balloon.h5 --image=<file name or URL>” was use to run the program.
INDEPENDENT STUDY CONTRACT PROJECTS

Note: Enrolment is subject to approval by the course convener

SECTION A (Students and Supervisors)

UniID: 15527568

SURNAME: XU

FIRST NAMES: TAO

PROJECT SUPERVISOR (may be external): Dr W M (Steve) Lee

FORMAL SUPERVISOR (if different must be an RSES academic):

COURSE CODE, TITLE AND UNITS: COMP8755 Individual Computing Project

COMMENCING SEMESTER: S2

YEAR: 2019 Two-semester project (12u courses only):

PROJECT TITLE:

Intelligent Microscopy: Applying Machine Learning to real-time biological dynamic cell tracking

LEARNING OBJECTIVES:

1. Collecting and pre-processing existing cell data (video, image) using customised video rate microscope
2. Using U-net or OpenCV to build a model to capture and process real-time biological images.
3. Training and testing model to ensure real-time selective microscopy imaging.

PROJECT DESCRIPTION:

Tracking biological cells is powerful tools that understand cell-cell interaction, cell-ECM interaction and also motility. Existing microscopy is time intensive due to manual selection of cells during imaging. In this project, we aim to develop a new microscopy imaging model that can intelligently select and image cells of interest.

To do that, we aim to conduct real-time microscopy using a specialized state of the art microscopy unit called PolySceo developed at the ANU in collaboration with TrendBio Scientific.

We will implement machine learning into a PolyScope software so that it can detect the specific cell tracking autonomously in real-time and record cellular division, cellular movement without any human intervention.

This will usher in the first real-time intelligent microscopy using PolyScope.

We will do the project in three steps:
1. Using existing data to build the model
2. Implementing model into software
3. Cell tracking application for live cells using our developed software
ASSESSMENT (as per the project course’s rules web page, with any differences noted below).

<table>
<thead>
<tr>
<th>Assessed project components:</th>
<th>% of mark</th>
<th>Due date</th>
<th>Evaluated by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report: style: research report, (e.g. research report, software description, ...)</td>
<td>45</td>
<td></td>
<td>(examiner)</td>
</tr>
<tr>
<td>Artefact: kind: software (e.g. software, user interface, robot, ...)</td>
<td>45</td>
<td></td>
<td>(supervisor)</td>
</tr>
<tr>
<td>Presentation:</td>
<td>10</td>
<td></td>
<td>(course convener)</td>
</tr>
</tbody>
</table>

MEETING DATES (IF KNOWN):

STUDENT DECLARATION: I agree to fulfil the above defined contract:

[Signature] 31/07/2019

SECTION B (Supervisor):

I am willing to supervise and support this project. I have checked the student’s academic record and believe this student can complete the project. I nominate the following examiner, and have obtained their consent to review the report (via signature below or attached email)

[Signature] 31/07/2019

Examiner:
Name: Miao Miao (CVPR), [Signature]
(Nominated examiners may be subject to change on request by the supervisor or course convener)

REQUIRED DEPARTMENT RESOURCES:

SECTION C (Course convener approval)

[Signature] Date

Research School of Computer Science

Form updated Jan 2018