

國立臺灣大學電機資訊學院生醫電子與資訊學研究所



碩士論文

Graduate Institute of Biomedical Electronics and Bioinformatics

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

免標註顯微鏡影像上以卷積神經網路預測粒線體結構

Mitochondrial Structure Prediction in Label-free
Microscopy Images Using Convolutional Neural Networks

許展銘

Chan-Min Hsu

指導教授：魏安祺 博士

Advisor: An-Chi Wei, Ph.D.

中華民國 109 年 7 月

July, 2020

Acknowledgments



書寫一篇論文，對於兩年前入學的我來說，看來是遙不可及的。但在兩年後的現在，看著自己所完成的內容，那堆疊在層層數據背後的實驗，似乎都有所回報了。回想當初踏入台大生醫電資所時的興奮、尋找研究方向時的迷惘、日復一日做實驗時的苦澀，是那麼的模糊卻又清晰，彷彿不曾存在卻有真實到刻骨銘心。

感謝生醫電資所給予的資源與環境，以及安排的學術活動，讓我在潛心研究之餘，能夠吸收來自不同領域的新知及養分，並且開拓眼界。

感謝我的指導教授魏安祺老師，在我的研究過程中給予許多指導與幫忙。在我剛入學的時候，讓我有足夠且自由的時間修課、探索專業；在我找不到研究方向的時候，引導我並適時的給予建議；在我受挫徬徨的時候，陪伴並且鼓勵著我。沒有她的指導與關切，我不會懂得如何進行研究、培養正確的研究精神，更不可能順利完成論文。

感謝張壯榮老師、劉彥良老師、何亦平老師，在忙碌之餘願意撥冗參與我的論文口試，並且給予許多建議。也特別感謝何亦平老師在我待在香港中文大學時，幫助我熟悉環境並且給予我研究上的支持。

我也要感謝一同在實驗室打拼的大家。謝謝張辰、文緯、祐德、姿樺在我有實驗上的問題時願意協助我；謝謝昱志、建勳，陪我參加各種休閒活動；謝謝紹庭，認識這麼久以來，不管在修課、參加研討會上，都幫助我很多，從你那學到的東西真的讓我獲益良多；也謝謝景詳、麒戎，為實驗室提供活力；特別謝謝羿儒，在我最忙碌的時候幫我培養細胞並拍照，也教了我很多實驗相關的知識，能夠完成論文有很大一部分都是妳的功勞。畢業後再回首，那些一起參加的活動、實驗的挫折、閒聊的人生，都會是我難忘的回憶。

最後，我要感謝我的父母。很抱歉在忙論文的這段時間很少聯絡，但也是因為有你們一路上的支持與信任，我才能義無反顧地朝自己夢想前進。我愛你們。

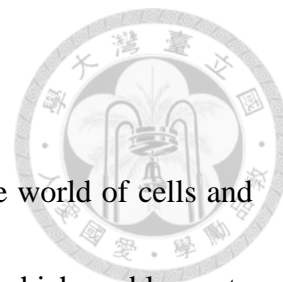


摘要

近年來顯微鏡技術的快速發展，讓科學家們能以更微觀的角度觀察細胞及細胞內的結構。在這之中，螢光顯微鏡因為能夠透過針對特定胞器進行螢光染色，藉此來觀察活細胞內的胞器活動，因而成為現今觀測活細胞時所主要使用的技術。然而透過雷射激發螢光的方式卻容易產生光漂白或是光毒性等問題，造成觀測上的困難。相比之下，單純的穿透光照射，雖然無法清晰看到細胞結構等細節，但便宜、不用染色的優點，讓它得以有不同的用途。本篇論文使用我們實驗室所拍攝的 AC16 心臟細胞穿透光及螢光影像，透過深度學習的方法並且採納 Allen Institute for Cell Science 發表的模型及方法來訓練，來實現三維及時間序列的螢光影像預測。相對於傳統的機器學習方法，卷積神經網路等深度學習方法近年來成功在影像辨識及切割上取得重大成果。因此，能夠透過訓練類似的神經網路模型，學習穿透光與螢光影像的相關性，最後成功從新的穿透光影像預測出其對應的螢光影像。在這裡，我們的實驗將著重在使用共軛焦顯微鏡及 Airyscan 激光共聚焦顯微鏡所拍攝的高解析粒線體影像及其 DNA 影像，進行不同條件（三維、時間序）的預測。總體來說，有了穿透光預測的螢光影像結果，將能有效縮短未來準備樣本的時間、增加細胞在螢光顯微鏡下可供觀測的時間，並讓研究者能更仔細的分析粒線體及粒線體 DNA 的形狀與動態。

關鍵字：粒線體結構、卷積神經網路、U-Net、顯微鏡影像預測、三維螢光影像、時間序列螢光影像

Abstract



Advancements in microscopic techniques allow insight into the world of cells and cellular structure. One such innovation is fluorescence microscopy, which enables us to analyze the subcellular structure of a living cell with the advantage of specific labeling. However, this technique comes with the potential problem of phototoxicity. The presence of advantage and disadvantage also holds for transmitted light microscopy (TL), which is a low-cost and label-free technology that nonetheless fails to easily distinguish targeted subcellular objects. In this thesis, we adopted the label-free method developed by the Allen Institute of Cell Science in using our TL microscopic images of cardiac myocyte-derived cell line AC16 to train and predict 3D (z-stack) and time-series fluorescence images. Convolutional neural networks (CNNs) have shown significant success in image recognition and segmentation compared with traditional machine learning methods. Based on a CNN-like U-Net architecture, the model can effectively predict fluorescence images from new TL input by learning the relationships between live-cell TL and fluorescence images for different kinds of subcellular structures. We specifically focused on building corresponding models of subcellular mitochondrial structures using the aforementioned CNN technology and compared the prediction results derived from confocal microscopic, Airyscan microscopic, z-stack, and time-series images. With the multi-model combined prediction, it is possible to generate integrated images through

only TL input, reduce the time required for sample preparation, increase the time scales to enable visualization and measurement, and understand the morphology and dynamics of mitochondria and mtDNA.



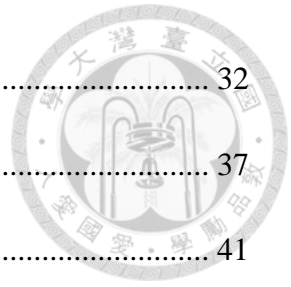
Keywords: mitochondrial structure; convolutional neural networks; U-net; microscope image prediction; 3D fluorescence images; time-series fluorescence images

Table of Contents



Acknowledgments	i
摘要	ii
Abstract.....	iii
Table of Contents	v
List of Figures.....	vii
List of Tables.....	ix
Chapter I: Introduction	1
Section 1-1: Background and Motivation	1
Section 1-2: Literature Review	5
Section 1-3: Specific Aims	9
Chapter II: Methods and Materials	12
Section 2-1: Cell Culture and Labeling	12
Section 2-2: Cell Imaging	12
Section 2-3: Data Preprocessing for Training and Evaluation.....	17
Section 2-4: Model Architecture.....	19
Section 2-5: Model Performance Analysis	23
Chapter III: Results.....	24
Section 3-1: Time-series Prediction.....	24
Section 3-2: Z-stack Prediction.....	28
Section 3-3: High-resolution Image Prediction (Without downscaling)	29

Section 3-4: Airyscan Prediction and Confocal Prediction	32
Section 3-5: Prediction from General Model.....	37
Chapter IV: Discussion.....	41
Section 4-1: Label-free Prediction on Mitochondria	41
Section 4-2: Result Analysis	42
Section 4-3: Experimental Difficulties	45
Section 4-4: Limitations.....	47
Chapter V: Conclusion and Future Work.....	50
Reference	51

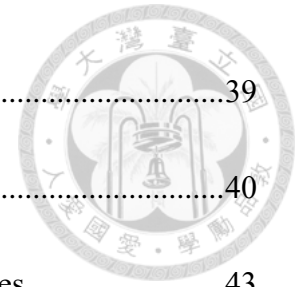


List of Figures



1.1 Prediction of mitochondria structure from bright-field images.....	4
1.2 The overall workflow of the experiment design.....	10
2.1 The Zeiss Microscope LSM800 with Airyscan	13
2.2 The diagram of original CNNs	19
2.3 The diagram of modified CNNs	21
3.1 Prediction in time-series (5-second interval).....	25
3.2 Prediction in time-series (15-minute interval).....	27
3.3 Prediction performance across different time intervals	28
3.4 Prediction in z-stack (0.29 μ m interval)	29
3.5 Prediction performance across different models and structures	29
3.6 High-resolution prediction in time-series	30
3.7 High-resolution prediction in z-stack	31
3.8 Prediction performance across different resolutions	31
3.9 Prediction in time-series with different imaging methods	33
3.10 Prediction in z-stack with different imaging methods.....	34
3.11 Prediction performance across different imaging methods	35
3.12 Summary results of methods used on this chapter	36
3.13 Prediction across the time-series models and the combined models.....	38

3.14 Prediction across the z-stack models and the combined models	39
3.15 Prediction performance across different model types	40
4.1 Prediction performance across different number of training images	43
4.2 Prediction performance across images of different time	44
4.3 Time-series prediction performance across different model types	44
4.4 Prediction performance across different patch sizes	45



List of Tables



2.1 Imaging parameters for each experiment	13
2.2 3D live cell imaging data.....	17
2.3 Hyperparameters used in the experiments.....	22
3.1 Performance metrics	35

Chapter I: Introduction

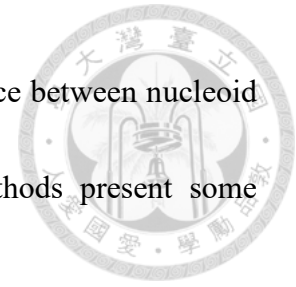


Section 1-1: Background and Motivation


Mitochondria are dynamic organelles that regulate cellular energy, function, cell fate, and survival[1]. The mitochondrial morphology is corresponded with the metabolic state and varies during different tasks such as cell division and differentiation[2]. In addition, mitochondria morphology and function can reflect the cellular health state, which makes them an excellent object for scientists to evaluate cell health. Common diseases such as neurodegenerative diseases, cancer and diabetes, all observed in changes of mitochondrial morphology and dynamics[3]–[5]. Researchers have also used the mitochondria phenotypes to distinguish tumor types and classify them[6], [7].

Because of the importance lying in mitochondria, it will be a cinch to further analyze and quantify changes in their dynamic behavior and morphology. Traditional cell imaging used fixed samples to conduct experiments, which cannot help researchers reveal the dynamism of mitochondria[8]. Hence, the researchers started to use live cell imaging to observe the details of cellular substructure decades ago. Such techniques are being improved over time to further help understand the structure of these organelles. For example, in the paper by Jevtic et al.[9], structured illumination microscopy (SIM) was used to track mitochondrial nucleoids in SYBR Gold labeled live cells. By comparing

each frame of SIM time-lapse images, they can calculate the distance between nucleoid positions with algorithms. However, most live-cell imaging methods present some restrictions.



The technique of fluorescence microscopy has been adopted in live-cell imaging in the past decade, and indeed achieves big progress such as performing confocal time-lapse and z-stack imaging in live cells[10]. Fluorescence microscopy has advantages of resolving subcellular structure (e.g. mitochondria, nucleus, etc.) in living cells by specific labeling, but it needs extra technical instrumentation and the sample preparation is time-consuming. In addition, the challenge of photobleaching and phototoxicity emerged when it comes to both z-stack and time-lapse imaging, creating a tradeoff between quality (image resolution) and timescales available for live cells[11]–[13]. In contrast, transmitted light microscopy (TL), e.g., bright-field, DIC, requires no labeling, which can significantly reduce phototoxicity[14]. The procedure of sample preparation will be simplified. Also, without the demand for dyes, the experiment cost will decrease. For example, when conducting an experiment to study Mitochondria and DNA, without the need for the corresponding dyes (TMRM and SYBRTM Gold), one can save up to 500 USD.



Although we can get information about cellular structure from TL images, such data lack distinct contrast compared to fluorescence images. Facing these issues, scientists have come up with solutions from the perspective of computing, more specifically, deep learning. As we know, huge advances in deep learning recently show its potential to achieve significant success in image processing. Convolutional neural networks (CNNs), one branch of deep learning, can learn non-linear relationships between source images and target images, resulting in considerably improved performance for computer vision tasks (classification, segmentation). Hence, with the help of CNNs, a method combining the relative low-cost TL images with clear fluorescence images would be a useful tool for observing subcellular structures.

To implement such tasks, Allen Institute for Cell Science and Google have previously constructed the convolutional neural networks to successfully predict fluorescence images from TL images[15], [16]. In their works, they trained the CNN model with unlabeled images (TL images) and fluorescent-label images, trying to learn relationships between each corresponding image. Consequently, the trained model can eventually predict fluorescence images from new unlabeled images. In that way, the phototoxicity and photobleaching problems can be avoided[15]. Moreover, without labeling, they can lower the cost of sample preparation and simplify the procedure.



Though the prediction works done by Allen Institute for Cell Science and Google have addressed the 3D cell imaging thoroughly, their model cannot fit with our own AC16 data (Fig. 1.1). The reason is few of their works are focus on the mitochondria imaging, especially in time-series images that could capture mitochondrial dynamics

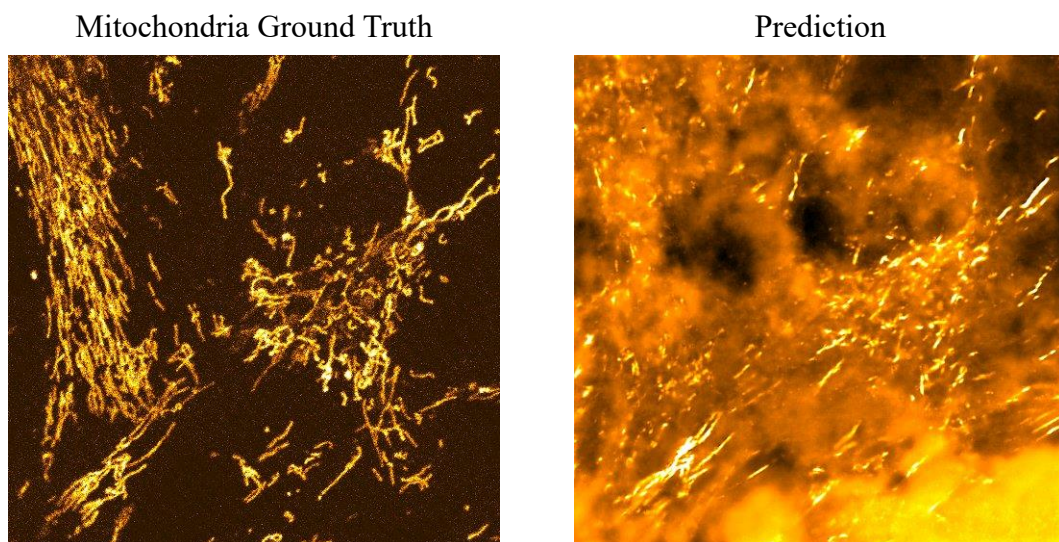


Figure 1.1: Prediction of mitochondria structure from bright-field images of AC16 cells by the “Mitochondria” model trained by Allen Institute for Cell Science[15].

As mentioned previously, we can gain many insights by studying the morphology and dynamics of mitochondria. Therefore, in this research, we adopted and modified the label-free U-Net method[15], [17] published by Allen Institute for Cell Science to train and predict z-stack and time-series mitochondria-related fluorescent images from the TL images. Using both confocal[18] and Airyscan[19] technology, we have improved the


overall performance of the CNN models to the AC16 cardiac cell line images.



Section 1-2: Literature Review

Fully Convolutional Networks

Fully convolutional networks (FCNs)[20] generally consist of multiple layers which are all convolutional layers, and are often used as semantic segmentation tasks. The term “semantic segmentation” is a pixel-wise prediction for the reason that each pixel in an image is classified according to its class. In the classical image classification task of CNN, an input image is downsized via several convolutional layers and ends with fully connected layers, and the output will be one of the predicted labels for that input image. Unlike CNN, FCN replaces the fully connected layers with other convolutional layers, resulting in images in smaller sizes. The upsampling technique is often used in this case. Since the output images will be smaller than the input images due to convolution, to achieve the goal of semantic segmentation, FCN implements the deconvolutional layers (transposed layers) to increase the size of the result images. However, with only upsampling from small size images (e.g., 2x2), the final heatmap will be rough because the model loses spatial information during the convolution process. To solve the issue, FCN fuses the output images with the previous output images from shallower layers.



FCN is one of the most efficient machine learning tools for semantic segmentation in the last decade. There are many extended applications and modifications based on FCN all gaining huge success. For example, R-FCN[21] demonstrated a modified FCN that combines pixel classification and object detection. Our current study uses the U-Net[17] architecture which was based on FCN for image prediction. The underlying U-Net system will be elaborated in the next section.

U-Net

U-Net[17] provides powerful semantic segmentation with two parts of convolutional layers, including a contracting path to capture context and an expanding path to localize information of each pixel. Similar to FCN structure, contracting path and expanding path represent traditional convolutional layers and deconvolutional layers, respectively. There are two primary differences between them: (1) In fusing step, unlike FCN which simply adds the value of two outputs together, U-Net concatenates the channels of each output, creating a larger size in channel-axis. (2) In most cases of biomedical images, there is very little training data available. As a result, using data augmentation as preprocess is necessary[22]. By applying elastic deformations to original training data, the network can learn invariance which is especially important in biomedical image segmentation since the most common variation in tissues and cells is deformation.




A general CNN focuses on image classification where input is an image and output is a label, and which is difficult to apply in most biomedical cases. The main purpose of biomedical images is to not only distinguish whether there is a target but also localize the area, which U-Net can easily apply. Hence, most biomedical images, e.g., MRI, ultrasound, and microscope, take advantage of U-Net nowadays.

Label-free Prediction of Fluorescence images from TL images

In the paper by Ounkomol et al. (Allen Institute for Cell Science)[15], a label-free method built from CNN-based U-Net architecture was introduced. The model can predict 3D fluorescent images directly from transmitted light images. Their prediction models learn each relationship between TL and cell images for several subcellular structures (cell membrane, nuclear envelope, DNA, mitochondria, microtubules, actin, ER, etc.). After training, a single 3D TL input can then be applied to multiple subcellular structure models, generating integrated images. In other sets of experiments, their method also can predict classical 2D immunofluorescence (IF) images from electron micrographs (EM).


To further examine the performance of the architecture under different imaging environments, they trained and tested the model with EM images and IF images from the



same samples, but by using two different microscopes. Thus, these two sets of images are not spatially aligned. It turned out that the model can successfully register an EM image to a target IF image, suggesting this label-free tool can be adopted in diverse imaging modalities.

In addition, they used the model trained on static images to predict fluorescence time-series images. By applying their TL-to-fluorescence models to TL 3D time-series at 5-minute intervals, the model can not only visualize subcellular structures but also recognize the dynamics of mitotic events. As reported in the previous researches[13], [23], phototoxicity occurs during a long period of acquisition of multi-label live cell fluorescence images, which indicates serious challenges in obtaining this type of information. Hence, their method prevented such problems from happening and extended the duration on which cellular processes can be visualized.

In another recent paper published by Christiansen et al. from the Google team[16], they built neural network models similar to the one proposed by Allen Institute for Cell Science[15] to predict fluorescent labels in unlabeled images (TL or DIC) but using different approaches. First, their model has a multi-scale input, including small and large length-scale paths. For the small length-scale path, the model processes explicit detail



near the center of the input. On the other hand, coarse information is processed via a large length-scale path in a broad region around the center. Such multi-scale architectures have been proof to be useful in computer vision works[24]. This approach helps the model better learn the spatial relationship between labeled and unlabeled data. Second, they trained the network with images of live cells treated with propidium iodide (PI), a dye that labels dead cells. The network accurately distinguished a single dead cell from a mass of live cells based on nuclear morphology. Last, they tried to predict specific subcellular structures in different plating processes. To their surprise, the network could perform well in conditions of low- and high- plating densities whereas the results are better under conditions of low-plating densities.

However, neither of these models conducts the prediction work on high-resolution microscope images. Moreover, although Allen Institute for Cell Science has dealt with time-series 3D images, they used static 3D z-stack images as training data. Therefore, to handle these issues, I adopted and modified the method published by Allen Institute to further look into the structures of mitochondria, along with high-resolution time-series and z-stack images.

Section 1-3: Specific Aims

The overall workflow of our experiment design is elaborated in Fig.1.2, with the following specific aims:

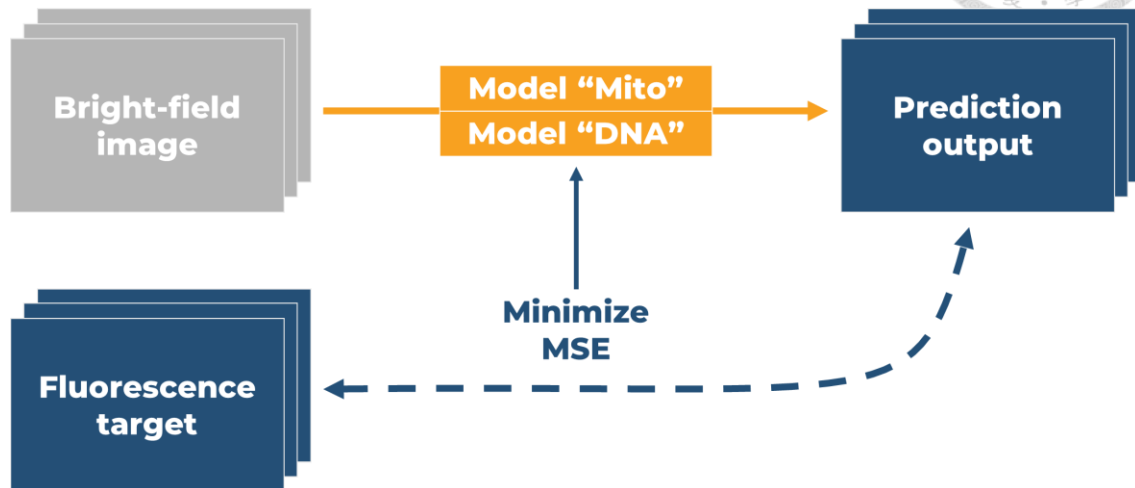



Figure 1.2: The overall workflow of the experiment design.

Aim1. High-Resolution Imaging

The high-resolution and high-contrast mitochondria images are acquired using a Zeiss confocal microscope (LSM800) with a 1.40-NA, 63x objective, and Airyscan option[19]. By optimizing labeling and image acquisition protocols, we could obtain fine details of the mitochondria images, allowing our networks to learn relationships between TL and fluorescent labels.

Aim2. Optimized U-Net Model for Mitochondrial Application

While acquiring input images in higher resolution, there will be more pixels in total



for the network to train. For example, the optimal size of an Airyscan image is around 1800x1800 pixels comparing to the confocal images of 500x500 pixels. To efficiently extract the information from the input, the original model[15] will be modified with deeper layers to preserve extra features inside the images. The improvement of the model and the proposed model architecture will be discussed in the next chapter.

Aim3. Specific Mitochondria Label Prediction with TL Input

The model trained by our TL microscopy images and fluorescence images has the ability to predict the mitochondria label from TL inputs, which do not require fluorescent dye labeling. This approach allows us to study morphology and dynamics of mitochondria without facing problems like photobleaching and phototoxicity, and enable us to increase the duration for imaging and monitoring mitochondrial structures.

Chapter II: Methods and Materials



Section 2-1: Cell Culture and Labeling

AC16 Human Cardiomyocyte Cell Line (AC16) was used in the z-stack and time-series microscopy image acquisition. The cells were seeded on a 30 mm glass-bottom plate. For AC16, the culture media was the same as imaging media, which was DMEM/F12 with 12.5% FBS and 1% antibiotic-antimycotic. The cell density was 250 to 500 thousand cells per plate. To stain the cells, AC16 was first incubated in the imaging media with 100 nM Tetramethylrhodamine, methyl ester (TMRM) for 10 to 15 min, and then with 2000x to 5000x SYBR GoldTM (Thermofisher) for 20 to 30 min. The cells were washed with PBS and fresh imaging media before imaging.

Section 2-2: Cell Imaging

All the cells were imaged on a Zeiss microscope LSM800 with ZEN Blue 2.6 software and with a 1.40-NA, 63x objective. The cells were imaged in the microscopy incubator with 5% CO₂ on the same day we stained them. For each training image, there are a total of three channels with corresponding 8-bit (Confocal) or 16-bit (Airyscan) data: transmitted light (bright-field), DNA and mitochondrial DNA labeled with SYBR Gold, mitochondria labeled with TMRM. The image size for all images is 78 μm * 78 μm, which contained 1 to 5 cells of AC16 per image. Table 2.1 shows the imaging parameters for

each imaging type.



Figure 2.1: The Zeiss Microscope LSM800 with Airyscan.


Imaging Type	Acquisition	488 nm laser	561 nm laser	Bit Depth
5-second interval	Con	2%, 660V	2%, 680V	8
15-minute interval	Airy + Con	0.2%, 700V	0.2%, 760V	16
1-minute interval	Airy + Con	0.8%, 700V	1%, 750V	16
30-second interval	Con	2%, 700V	1%, 760V	8
z-stack	Con	2%, 660V	2%, 630V	8
HR z-stack (Airyscan)	Airy + Con	1%, 730V	1%, 750V	16
HR z-stack (Confocal)	Con	2%, 700V	1%, 750V	8

HR: high-resolution; Airy: Airyscan; Con: confocal

Table 2.1: Imaging parameters for each experiment. Note that some acquisitions combined two methods (Airyscan and confocal, Second column). The 488 nm laser was for SYBR Gold (Third column). The 561 nm laser was for TMRM (Fourth column). The upper part is time-series experiments and the lower part is z-stack experiments.

Time-series


In the 5-second interval acquisitions, each set was imaged for a total of about 5 min



using only confocal microscopy. The acquisition settings for each channel were: 488 nm laser at 2.00% of 10mW, with detector gain 660 V for SYBR Gold and 300 V for bright-field; 561 nm laser at 2.00% of 10mM, with detector gain 680 V for TMRM. The images were captured at a YX-resolution of 512 px * 512 px with a pixel scale of 0.152 $\mu\text{m}/\text{px}$. All time-series sets were composed of 64 frames with the same time interval.

In the long time interval acquisitions(15-minute), we choose the Airyscan detector [19] to implement the imaging process. The cell was imaged by Airyscan for two channels: SYBR Gold and TMRM, while the confocal microscope was only used for the bright-field channel. With longer time interval, the total duration for this experiment was about 16 h for each set (total 64 frames). The acquisition settings for each channel were: 488 nm laser at 0.2% of 10mW, with detector gain 700 V for SYBR Gold; 561 nm laser at 0.2% of 10mM, with detector gain 760 V for TMRM; 488nm laser at 2.00% of 10mW, with detector gain 300 V for bright-field. The images were also captured at a YX-resolution of 512 px * 512 px with a pixel scale of 0.152 $\mu\text{m}/\text{px}$. All time-series sets were composed of 64 frames with same time interval.


As for high-resolution time-series images, we used Airyscan and confocal detector to conduct the experiment. Since in high-resolution condition, the exposure time of laser



is much longer, to avoid the phototoxicity, we reduced the time-series set of 64 frames to 32 frames. For Airyscan images, the cell was imaged by Airyscan for both SYBR Gold and TMRM channels and the bright-field images were obtained by switching to the confocal microscope. The duration for high-resolution time-series Airyscan experiment was around 32 min for each set of 1-minute interval (total 32 frames). The acquisition settings for each channel were: 488 nm laser at 0.8% of 10mW, with detector gain 700 V for SYBR Gold; 561 nm laser at 1.00% of 10mM, with detector gain 750 V for TMRM; 488nm laser at 2.00% of 10mW, with detector gain 300 V for bright-field. The high-resolution images were captured at a YX-resolution of 1834 px * 1834 px with a pixel scale of 0.043 $\mu\text{m}/\text{px}$. For confocal images, the cell was imaged by confocal detector for all three channels. The duration for high-resolution time-series confocal experiment was 15 min for each set of 30-second interval (32 frames). The acquisition settings for each channel were: 488nm laser at 2.00% of 10mW, with detector gain 700 V for SYBR Gold and 300 V for bright-field; 561 nm laser at 1.00% of 10mM, with detector gain 760 V for TMRM. The high-resolution confocal images were captured at a YX-resolution of 917 px * 917 px with a pixel scale of 0.085 $\mu\text{m}/\text{px}$

Z-stacks

In the 512 * 512 z-stack experiment, cells were imaged for up to 3 min on confocal



microscope for 64 slices. The acquisition settings for each channel were: 488 nm laser at 2.00% of 10mW, with detector gain 660 V for SYBR Gold and 300 V for bright-field; 561 nm laser at 2.00% of 10mM, with detector gain 630 V for TMRM. The images were captured at a YX-resolution of 512 px * 512 px with a pixel scale of 0.152 $\mu\text{m}/\text{px}$. All z-stacks were composed of 64 slices with an interval of 0.100 μm .


For the high-resolution z-stack experiment, cells were imaged for up to 1.5 min on the confocal microscope and up to 15 min on the Airyscan method. The acquisition settings of the Airyscan method for each channel were: 488 nm laser at 1.00% of 10mW, with detector gain 730 V for SYBR Gold; 561 nm laser at 1.00% of 10mM, with detector gain 750 V for TMRM; 488nm laser at 2.00% of 10mW, with detector gain 300 V for bright-field. The acquisition settings of the confocal method for each channel were: 488nm laser at 2.00% of 10mW, with detector gain 700 V for SYBR Gold and 300 V for bright-field; 561nm laser at 1.00% of 10mM, with detector gain 750 V for TMRM. For confocal microscope z-stack images, the pixel size is 917 px * 917 px with a pixel scale of 0.085 $\mu\text{m}/\text{px}$; For Airyscan z-stack images, the pixel size is 1834 px * 1834 px with a pixel scale of 0.043 $\mu\text{m}/\text{px}$. Due to the phototoxicity problem in high-resolution images, all z-stacks were composed of 32 slices with an interval of 0.150 μm .

Section 2-3: Data Preprocessing for Training and Evaluation

Structure (Acquisition)	Model Name	3 rd dimension	Number of data (train/test)	Resolution	Number of slices (frames)
Mitochondria (Airyscan)	A ₁	Time-series (15 min)	90/30	512 * 512	64
DNA (Airyscan)	A ₂	Time-series (15 min)	90/30	512 * 512	64
Mitochondria (Confocal)	B ₁	Time-series (5 sec)	150/50	512 * 512	64
DNA (Confocal)	B ₂	Time-series (5 sec)	150/50	512 * 512	64
Mitochondria (Airyscan)	C ₁	Time-series (1 min)	47/20	1834 * 1834	32
DNA (Airyscan)	C ₂	Time-series (1 min)	47/20	1834 * 1834	32
Mitochondria (Confocal)	D ₁	Time-series (30 sec)	60/15	917 * 917	32
DNA (Confocal)	D ₂	Time-series (30 sec)	60/15	917 * 917	32
Mitochondria (Confocal)	E ₁	Z-stack (0.100 μ m)	36/14	512 * 512	64
DNA (Confocal)	E ₂	Z-stack (0.100 μ m)	36/14	512 * 512	64
Mitochondria (Airyscan)	F ₁	Z-stack (0.150 μ m)	47/18	1834 * 1834	32
DNA (Airyscan)	F ₂	Z-stack (0.150 μ m)	47/18	1834 * 1834	32
Mitochondria (Airyscan)	G ₁	Z-stack (0.150 μ m)	47/18	917 * 917	32
DNA (Airyscan)	G ₂	Z-stack (0.150 μ m)	47/18	917 * 917	32
Mitochondria (Confocal)	H ₁	Z-stack (0.150 μ m)	46/17	917 * 917	32
DNA (Confocal)	H ₂	Z-stack (0.150 μ m)	46/17	917 * 917	32

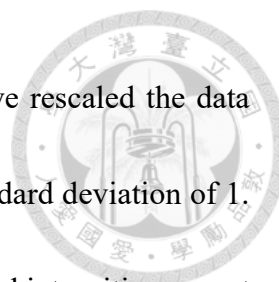
Table 2.2: 3D live cell imaging data used in this research. Each structure was imaged in Airyscan or Confocal acquisition (Leftmost column). Each 3D image was composed in the order of either “Length, width, height” or “Length, width, time” depended on the 3rd dimension (Third column). The total dataset was split into 75% training set and 25% testing set (Fourth column). The number of resolution presented length * width (Fifth column). The size of 3rd dimension is either 32 or 64 (Rightmost column). The abbreviated name of each model was shown in the table (Second column).

Table 2.1 shows the data and the details we used to train and evaluate. Our models are based on different types of data, including short-interval live cell time-lapse, long-



interval live cell time-lapse, 3D live cell z-stacks, and their corresponding pairs in high-resolution (512 * 512 vs 1834 * 1834). The data were split into 75% training sets and 25% testing sets, and then the training data were again split into 90% training sets and 10% validation sets. All multi-channel time-series, multi-channel z-stack data were obtained by our Zeiss LSM800 microscope. For Airyscan, we must first utilize “Airyscan process” function in ZEN Blue 2.6 to reconstruct the high-contrast image with 32 original images before doing subsequent procedures. Note that to compare the result of high-resolution Airyscan images (1834 px * 1834 px) with the result of confocal images (917 px * 917 px), Airyscan images were downscaled by a factor of two in YX dimension.

With the finished pairs, the model can be trained and evaluated to predict the tagged subcellular structure pixelwise. All images were converted to floating-point and resized via cubic interpolation for the purpose of having the same pixel scale in z-stacks ($0.1\mu\text{m} * 0.1\mu\text{m} * 0.1\mu\text{m}$ for original size and $0.15\mu\text{m} * 0.15\mu\text{m} * 0.15\mu\text{m}$ for high-resolution). Result images were 176 px * 176 px for original images, 396 px * 396 px for downscaled high-resolution images, or 512 px * 512 px for high-resolution images in Y and X respectively. For original z-stack and time-series, there will be 64 slices in the result; For high resolution ones, there will be 32 slices in the result. Last, to diminish the difference caused by illumination, pixel intensities of input and target images were improved by Z-



score normalization. In the process of the Z-score normalization, we rescaled the data intensities into a normal distribution with the mean of 0 and the standard deviation of 1. This step is necessary to make our ML task well defined since the pixel intensities are not measured in comparable absolute units.

Section 2-4: Model Architecture

Figure 2.2 shows the overall diagram of the convolutional neural networks model.

We adopt the method published by Allen Institute of Cell Science which is based on the U-Net architecture [17]. The modified U-Net model proposed in this thesis is shown in

Figure 2.3.

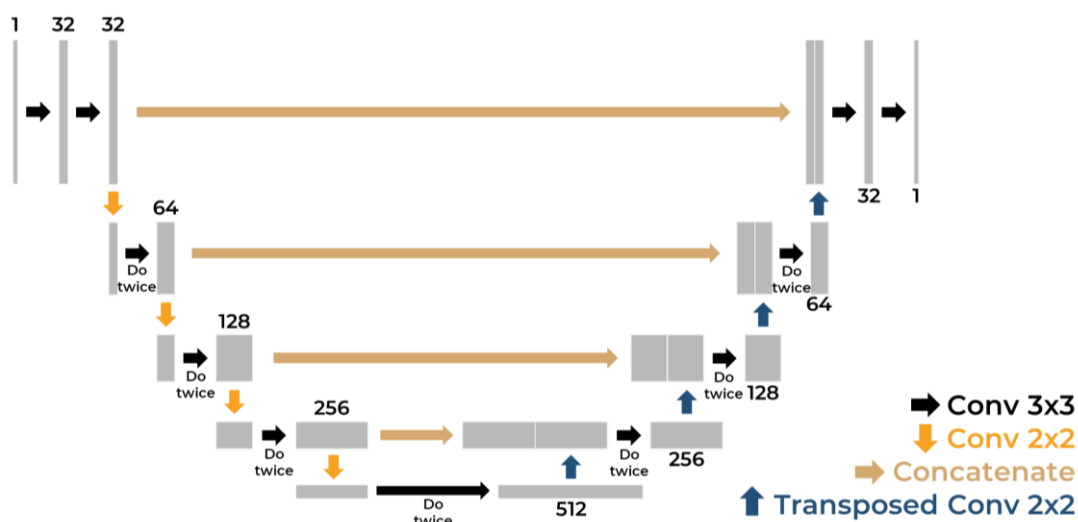


Figure 2.2: The original diagram of CNNs. Note that each layer consists of convolutional layer, batch normalization and ReLU function except the “concatenate” arrow and the rightmost arrow. Patch size of input is 32 * 64 * 64 in this model. Figured adapted from Ronneberger et al. Lecture Notes in Computer Science 234–241 (2015) [17] and Ounkomol et al. Nature Methods (2018) [15].



Our original model consists of different layers that perform three kinds of convolution functions: convolutional layers with a stride of 2 pixels (The orange arrows in Fig. 2.2), convolutional layers with a stride of 1 pixel (The brown arrows in Fig. 2.2), and deconvolutional layers with a stride of 2 pixels (The blue arrows in Fig. 2.2), each comes with a batch normalization[25] and ReLU function. To keep the size of the output layers the same as input size, the horizontal arrows represent the convolution layers that are $3 * 3$ pixels with a stride of 1-pixel using zero-padding on input. For the down arrows, they are $2 * 2$ pixels convolution layers with a stride of 2 pixels in order to halve the size of the output. Last, the up arrows represent the deconvolution (transposed convolution) layers that are $2 * 2$ pixels with a stride of 2 pixels for the purpose of doubling the output size. Note that in the last layer of the model, there are no ReLU function and batch normalization. The number of output channels in each layer are indicated in the Figure 2.2.

As mentioned in Section 1-2, the purpose of the left-hand side of the U-Net is to extract features from the input data while the right-hand side stands for reconstructing images from the features of the input. Since in every layer we would perform batch normalization[25] and ReLU function right after the convolution function, the output of



convolutional layer is considered reliable in pixel intensities.

To further improve the performance of high-resolution images, the bigger patch size is needed. Hence, we modified the original model by adding more convolutional layers, creating a deeper neural network which can extract more features from input data. As shown in Figure 2.3, there is one more arrow that goes down, meaning the model will keep doing convolution to halve the size. Therefore, with such fine details, we can deal with a bigger patch size without losing information.

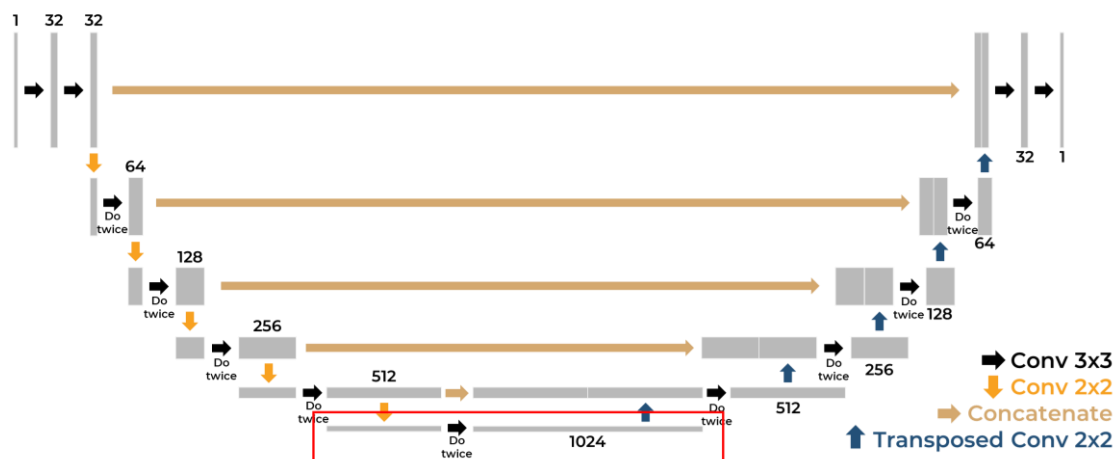


Figure 2.3: The diagram of modified CNNs. The difference between this model and the original one is highlighted in red rectangle. Patch size of input is $32 * 128 * 128$ in this model.

Our model was initially trained on a GeForce GTX 1080Ti with 12 GB RAM, with the patch size of $64 \text{ px} * 64 \text{ px} * 32 \text{ px}$. As patch size increases, we trained the model on several Tesla V100s with 32 GB RAM, using data parallelism as well. The GPU device



Tesla V100 was provided by Taiwan Computing Cloud (TWCC). For the details of TWCC, please refer to the TWCC website (<https://www.twcc.ai/>). All the patches were randomly subsampled across all training images. The training procedure updated its parameters via stochastic gradient descent to minimize the mean squared error. With a learning rate of 0.001 for different batch iterations (depends on the size of training data), we used the Adam optimizer[26] as an optimization method. Choosing a batch size of 24 for our 3D model, the total training progress took about 5 hours in our workstation and 4 hours in TWCC for 10000 iterations and so on. For prediction tasks, it takes 1 second for a single 3D image (XYZ or XYT) whose input pixel size is 512 px * 512 px, and 5 second for a single 3D image whose input pixel size is 917 px * 917 px. Both original and modified model training pipelines were implemented in Python, using the PyTorch package[27]. Table 2.3 shows the detail of hyperparameters for each corresponding model.

Model	Iteration Number	Patch size (ZXY or TXY)	Architecture Depth
A _{1,2} (Airyscan, 15min)	20000	32 * 64 * 64	4
B _{1,2} (Confocal, 5sec)	20000	32 * 64 * 64	4
C _{1,2} (Airyscan, 1min, HR)	10000	32 * 128 * 128	5
D _{1,2} (Confocal, 30sec, HR)	10000	32 * 64 * 64	4
E _{1,2} (Confocal, z-stack)	10000	32 * 64 * 64	4
F _{1,2} (Airyscan, z-stack, HR)	10000	32 * 128 * 128	5
G _{1,2} (Airyscan, z-stack, DS)	10000	32 * 64 * 64	4
H _{1,2} (Confocal, z-stack, HR)	10000	32 * 64 * 64	4

HR: high-resolution; DS: downscaling
A₁, B₁...H₁: Mitochondria model; A₂, B₂...H₂: DNA model

Table 2.3: Hyperparameters used in our experiment. The iteration numbers were depended on the number of training images (Second column). The dimensions of the patch were either ZXY or TXY (Third column). Architecture Depth referred to the different models (4: original model from Allen Institute, 5: modified model). The upper part is time-series models and the lower part is z-stack models. The abbreviation of each model is shown in Table 2.2.

Section 2-5: Model Performance Analysis

For our model performance, we choose the Pearson correlation coefficient to quantify the accuracy:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (2.1)$$

Here y stands for the pixel intensities of the model's prediction (output), and x stands for the ground truth test images. The closer x is to y , the r value will become bigger. It is intuitive that as the size of the signal increases, the size of random fluctuations will also increase, which consequently leads to degrading in model performance. For example, in some cases of the high-resolution image experiment, the model itself cannot classify between background noise and labeled target. Hence, the result may be good in human observation but bad in Pearson correlation coefficient.

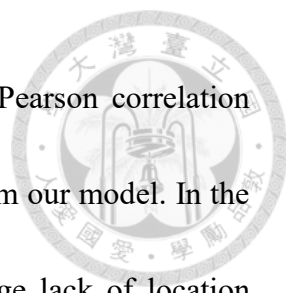
Chapter III: Results



In this chapter, we will present the predictions and the performance from TL images using different models. In Section 3-1, we compared the short interval time-series predictions of mitochondria and DNA with the long interval predictions. In Section 3-2, we presented the z-stack predictions using our model and compared the performance with the pretrained model published by Allen Institute for Cell Science. In Section 3-3, we compared the high-resolution (1834 *1834) predictions with the normal-resolution (512 *512) predictions. In Section 3-4, we compared the performance of the models trained by images acquired using different methods: Airyscan microscopy and confocal microscopy. Finally, in Section 3-5, we constructed a general model that was trained with both time-series and z-stack images.

Section 3-1: Time-series Prediction

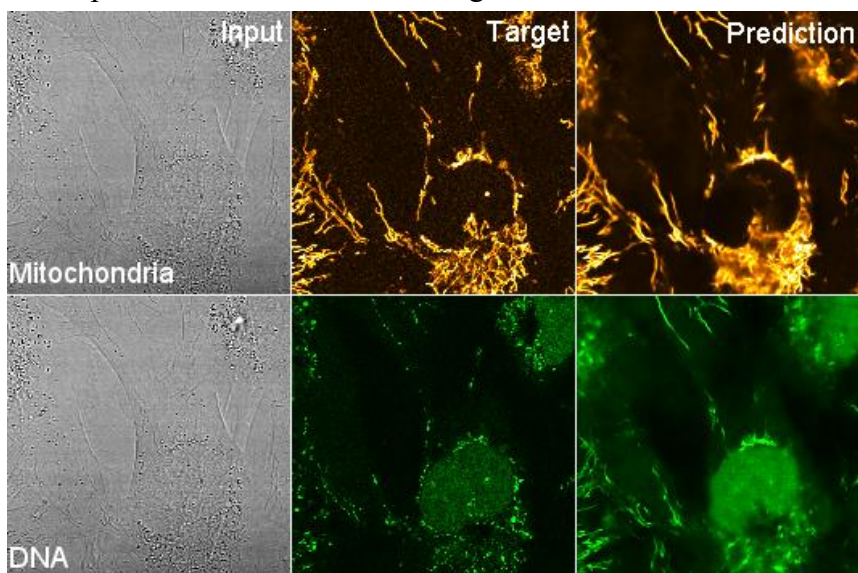
Fig. 3.1 and Fig. 3.2 show that the model could predict mitochondria and DNA from the transmitted light images. In the mitochondrial structure predictions, the model successfully classified between tubular mitochondria and background. But mitochondria network clustered around nucleus are hardly classified due to the overlap of structure in transmitted light. On the other hand, DNA prediction was mostly overlapped with mitochondria except regions of nucleus. However, the structure of mtDNA is not as clear



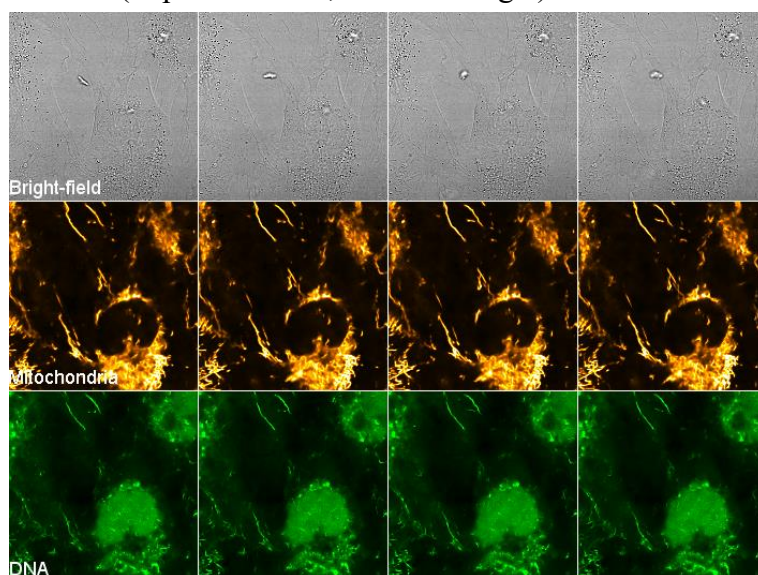
as shown in the fluorescence-labeled images. Fig. 3.3 shows the Pearson correlation coefficient (r) between ground truth (target) and predicted image from our model. In the experiment of long intervals, slices between each time-series image lack of location relationships, resulting in low performance.

5-second Interval

(a) Fluorescence prediction from transmitted light



(b) Time-series results (Top: Prediction; Bottom: Target)



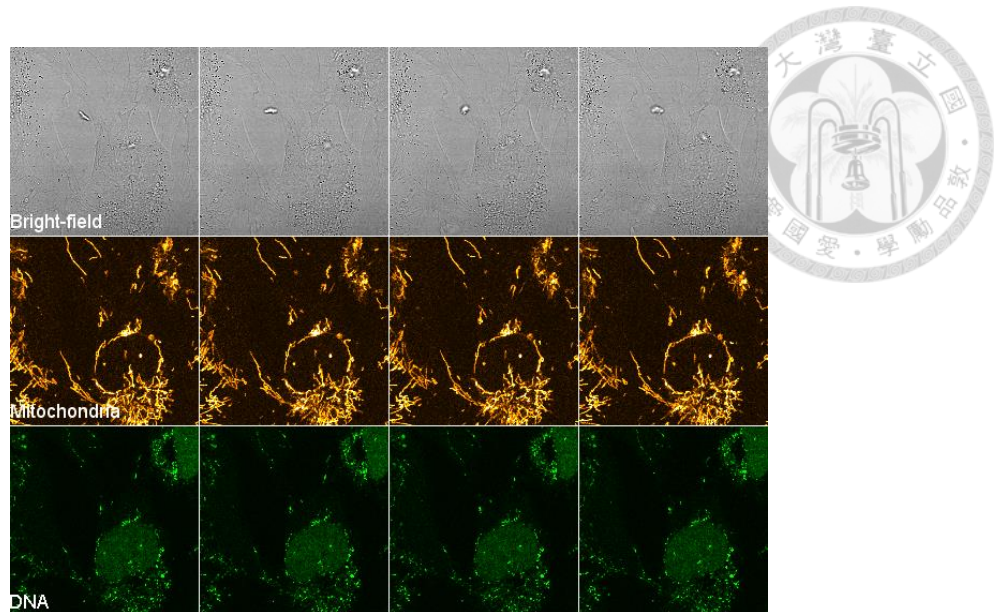
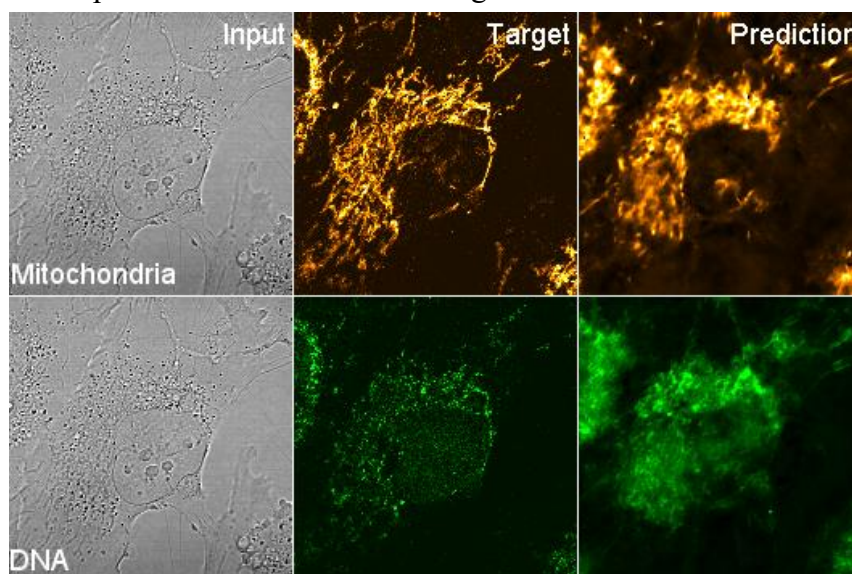


Figure 3.1: (a) Fluorescence prediction from transmitted light (5-second interval). The left column is the same single t-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images. (b) Time-series Prediction of mitochondria and DNA from transmitted light (top) and the ground truth of them (bottom). The interval between each pair is 5-second.

15-minute Interval

(a) Fluorescence prediction from transmitted light



(b) Time-series results (Top: Prediction; Bottom: Target)

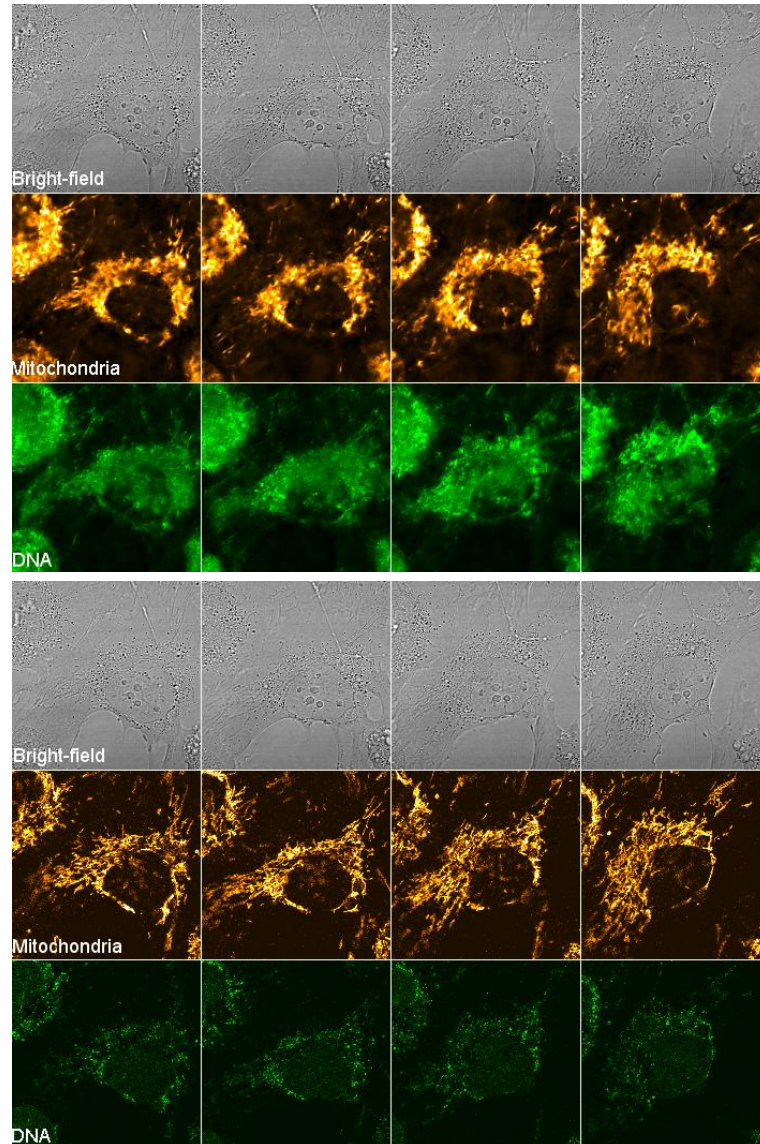


Figure 3.2: (a) Fluorescence prediction from transmitted light (15-minute interval). The left column is the same single t-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images. (b) Time-series Prediction of mitochondria and DNA from transmitted light (top) and the ground truth of them (bottom). The interval between each pair is 15-minute.

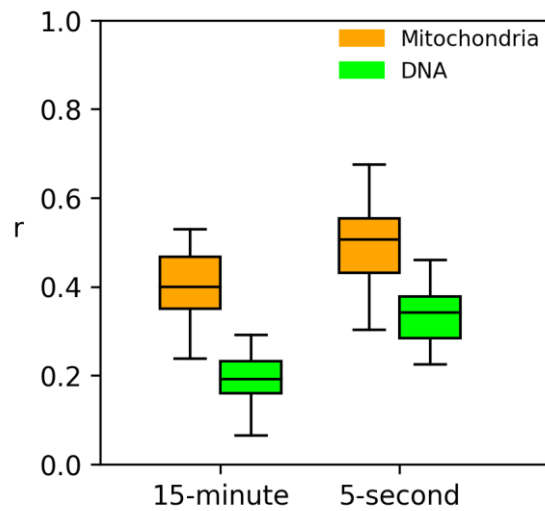


Figure 3.3: Prediction performance across different time intervals of time-series image acquisitions.

Section 3-2: Z-stack Prediction

Fig. 3.4 shows that the results were successfully predicted from transmitted light images. Compared with previous works done by Allen Institute for Cell Science[15], our z-stack prediction focused on mitochondria and DNA with higher magnification. Although the prediction performance of our model was slightly lower than theirs (Fig. 3.5a), our results provided more fine details about mitochondria structure under such magnification (1.40-NA, 63x objective in our confocal microscope vs 1.25-NA, 100x objective in spinning disk microscope from Allen Institute). It is noted that under such magnification, the noise and the dust will affect the performance significantly, resulting in bigger error bar (Fig. 3.5a).

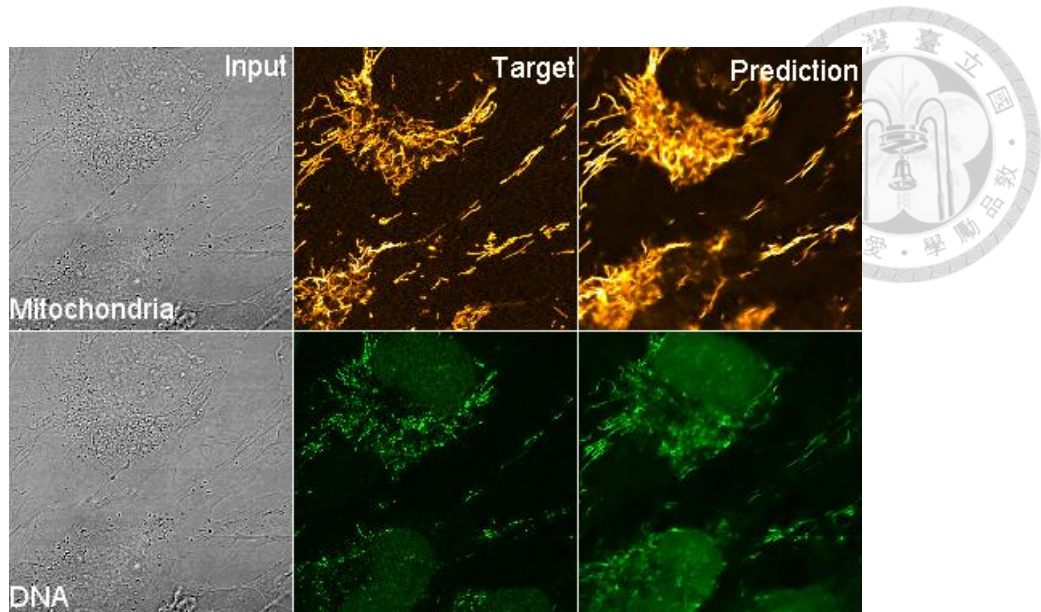


Figure 3.4: Fluorescence prediction from transmitted light (z-slice interval of $0.29\mu\text{m}$). The left column is the same single z-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images.

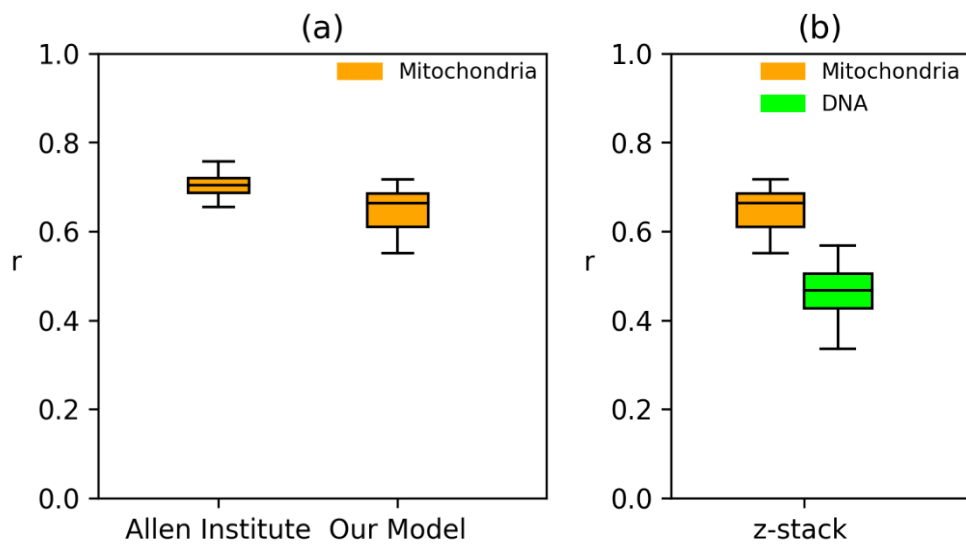


Figure 3.5: (a) Prediction performance on mitochondria across different models. (b) Prediction performance on different subcellular structures.

Section 3-3: High-resolution Image Prediction (Without downscaling)

Both Fig. 3.6 and Fig. 3.7 shows that the model could predict the subcellular structures from transmitted light images. However, the results were blurry because the background noise was also magnified under high-resolution TL images. Since the deep neural network structure is susceptible to blur and noise distortion[28], such magnified noise will lead to poor performance of the model. Fig. 3.8 shows that the average Pearson correlation coefficient (r) of high-resolution decreased in both time-series and z-stack.

Time-series

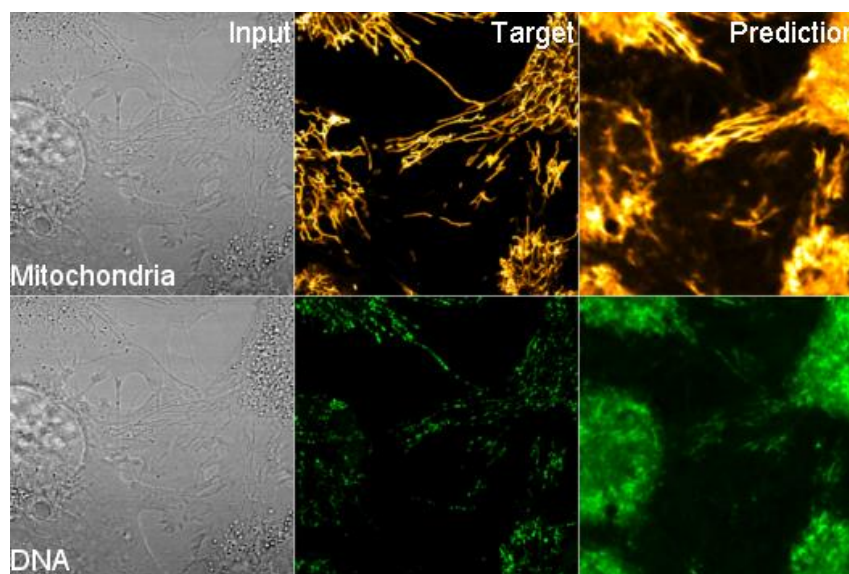


Figure 3.6: High-resolution fluorescence prediction from transmitted light (1-minute interval). The left column is the same single t-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images.

Z-stack

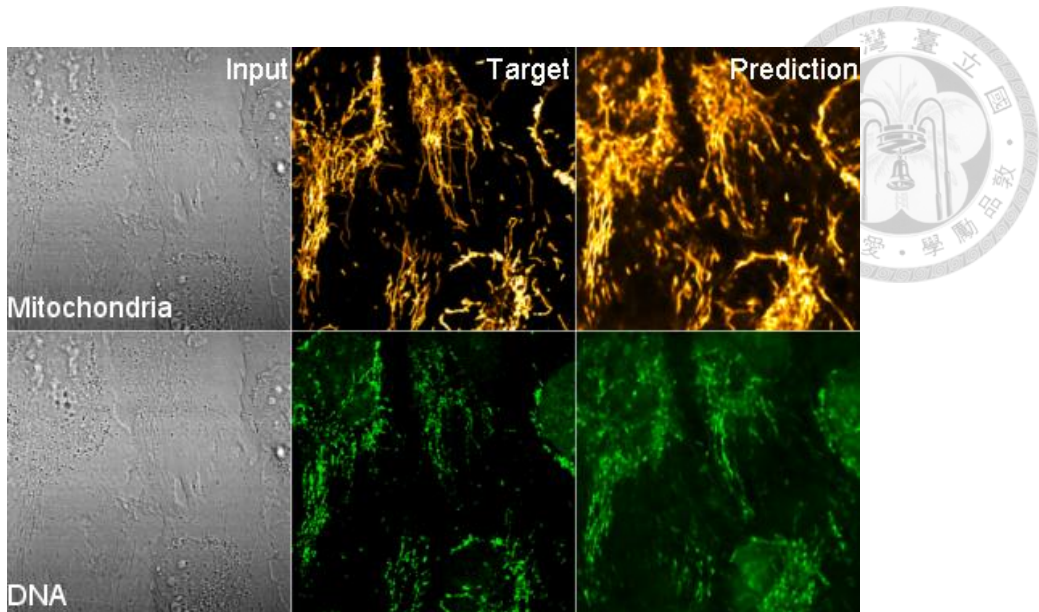


Figure 3.7: High resolution fluorescence prediction from transmitted light (z-slice interval of $0.15\mu\text{m}$). The left column is the same single z-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images.

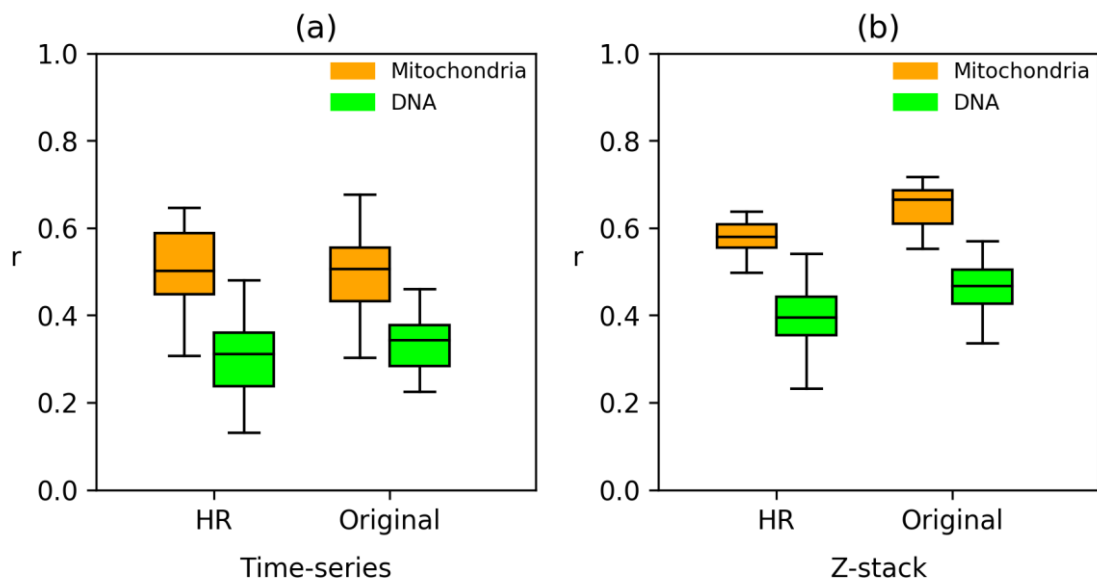
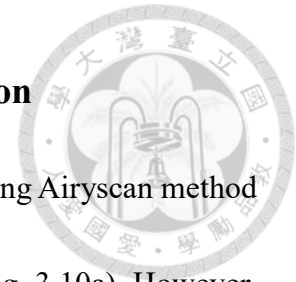


Figure 3.8: (a) Time-series prediction performance across high resolution (HR, $1834 * 1834$) and original resolution (Original, $512 * 512$). (b) Z-stack prediction performance across HR and original resolution.

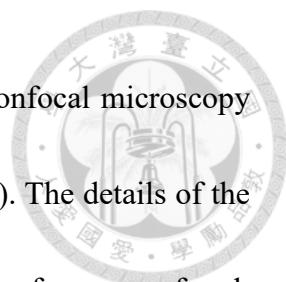
Section 3-4: Airyscan Prediction and Confocal Prediction



In both time-series and z-stack, the model trained on images using Airyscan method could predict mitochondria structure with fewer errors (Fig. 3.9a, Fig. 3.10a). However, the performance on DNA structure (Fig. 3.9a, Fig. 3.10a) was poor since Airyscan method will reduce the intensity of nucleus whose intensities were already low. The intensity of mtDNA remained the same or increased in the meantime. Consequently, this process will result in irrelevant pairs between TL input images and fluorescence target images.

In contrast, Fig. 3.9b and Fig. 3.10b show that the model trained on images using confocal method could predict mitochondria and DNA structures with higher accuracy. Because the images acquired using confocal microscopy provide high-contrast and relatively high-resolution without any imaging process function, such methods ensured the high-relevant relationship between TL input images and fluorescence target images. Fig. 3.11 shows that in both time-series and z-stack, the Pearson correlation coefficient (r) of the confocal imaging method was better than that of the Airyscan imaging method.

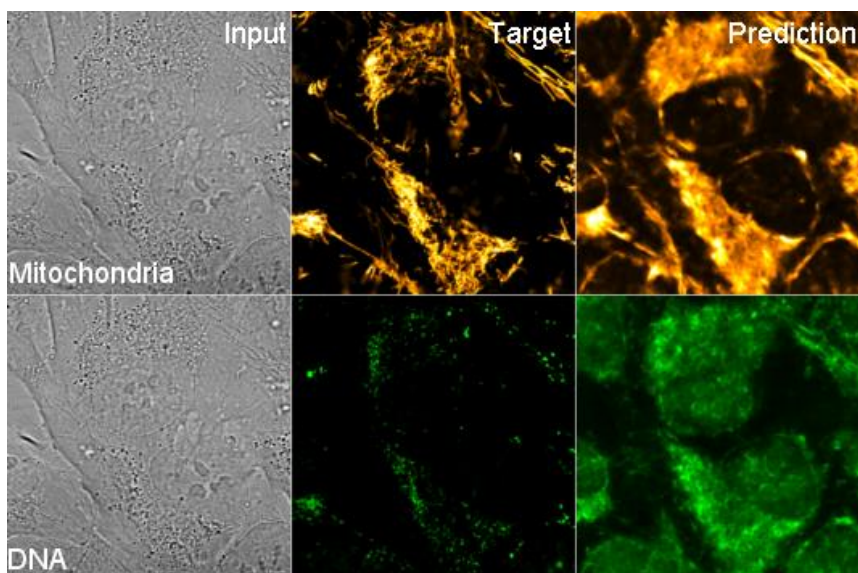
Table 3.1 shows the overall performance metrics for the models mentioned in this chapter. The performance metrics were composed of the Pearson correlation coefficient (r) for each subcellular structure, mitochondria and DNA. Compared with other models,



the model trained with the images acquired using high-resolution confocal microscopy has superior performance in both time-series ($D_{1,2}$) and z-stack ($H_{1,2}$). The details of the Pearson correlation coefficient as shown in Figure 3.12, includes the performance of each testing set in the model of $H_{1,2}$.

Time-series

(a) Airyscan



(b) Confocal

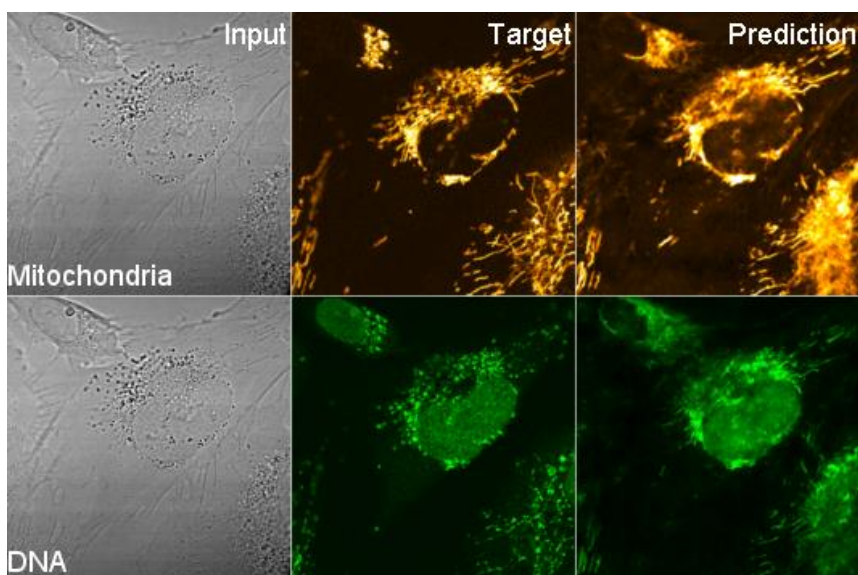
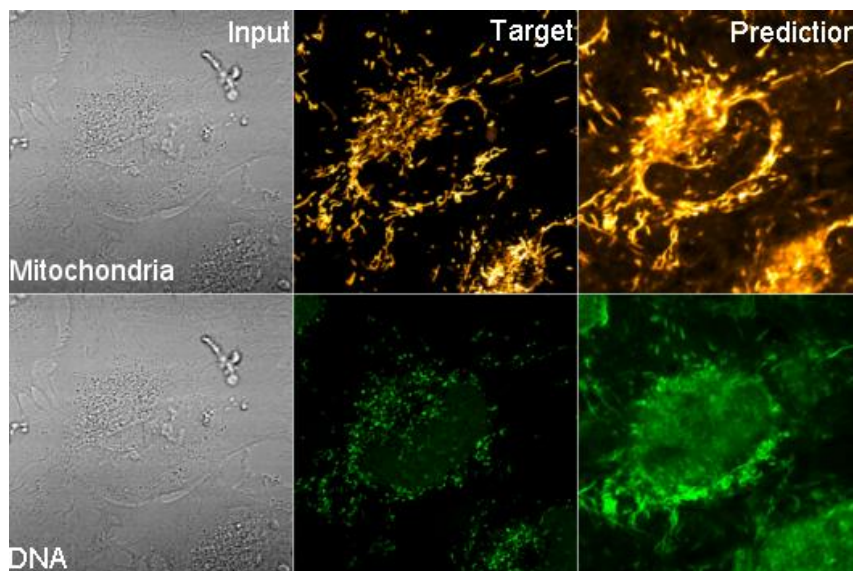


Figure 3.9: (a) Time-series fluorescence prediction from transmitted light using Airyscan. (b) Time-series fluorescence prediction from transmitted light using confocal microscopy. In both (a) and (b), the left column is the same single t-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images.

Z-stack

(a) Airyscan



(b) Confocal

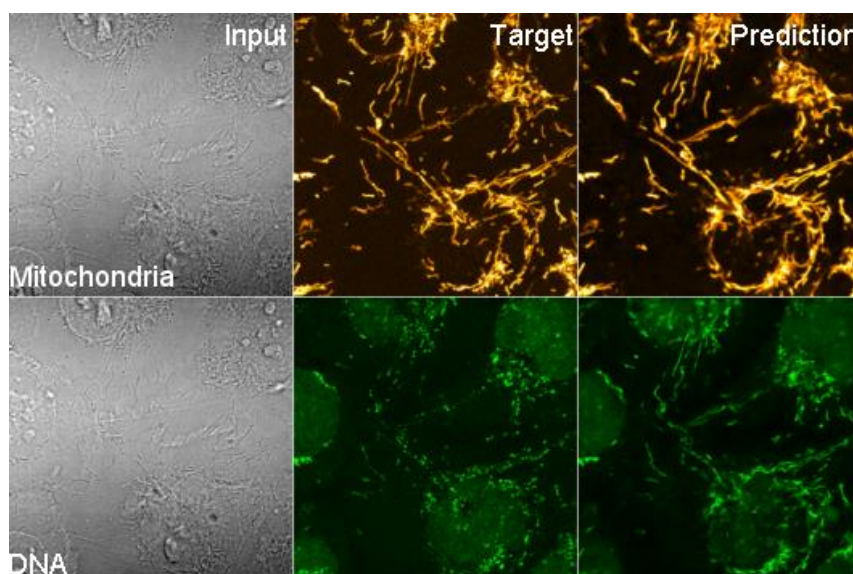


Figure 3.10: (a) Z-stack fluorescence prediction from transmitted light using Airyscan. (b)

Z-stack fluorescence prediction from transmitted light using confocal microscopy. In both (a) and (b), the left column is the same single t-slice of a 3D transmitted light input image. The middle column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The right column is images predicted by corresponding models (mitochondria model and DNA model). The example images are selected from a larger pool of test images.

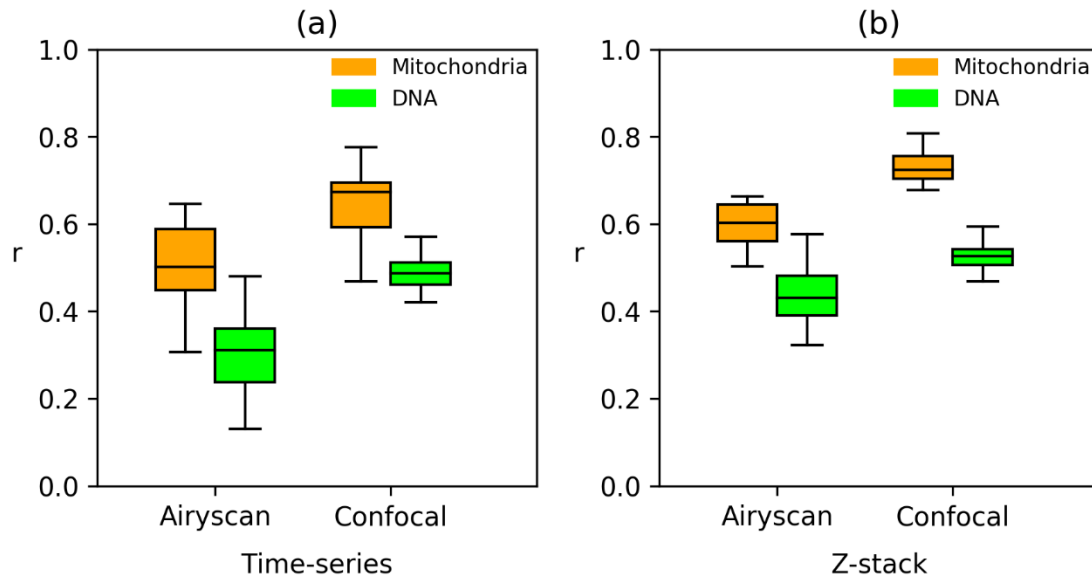


Figure 3.11: (a) Time-series prediction performance across different imaging methods. (b) Z-stack prediction performance across different imaging methods.

Dataset: LSM800 Microscope Images		
Models	Mitochondria	DNA
A _{1,2}	0.4020	0.2032
B _{1,2}	0.4956	0.3363
C _{1,2}	0.4989	0.2988
D _{1,2}	0.6431	0.4746
E _{1,2}	0.6488	0.4703
F _{1,2}	0.5633	0.3966
G _{1,2}	0.5879	0.4361
H _{1,2}	0.7312	0.5273

Table 3.1: Performance metrics for the considered models trained and tested using the microscope images as training and testing datasets, respectively. The upper part (A, B, C, D) is for the time-series prediction and the lower part (E, F, G, H) is for the z-stack prediction. All the model experiments are described in the text. The numbers in the table are the values of Pearson correlation coefficient. The abbreviation of each model is shown

in Table 2.2.

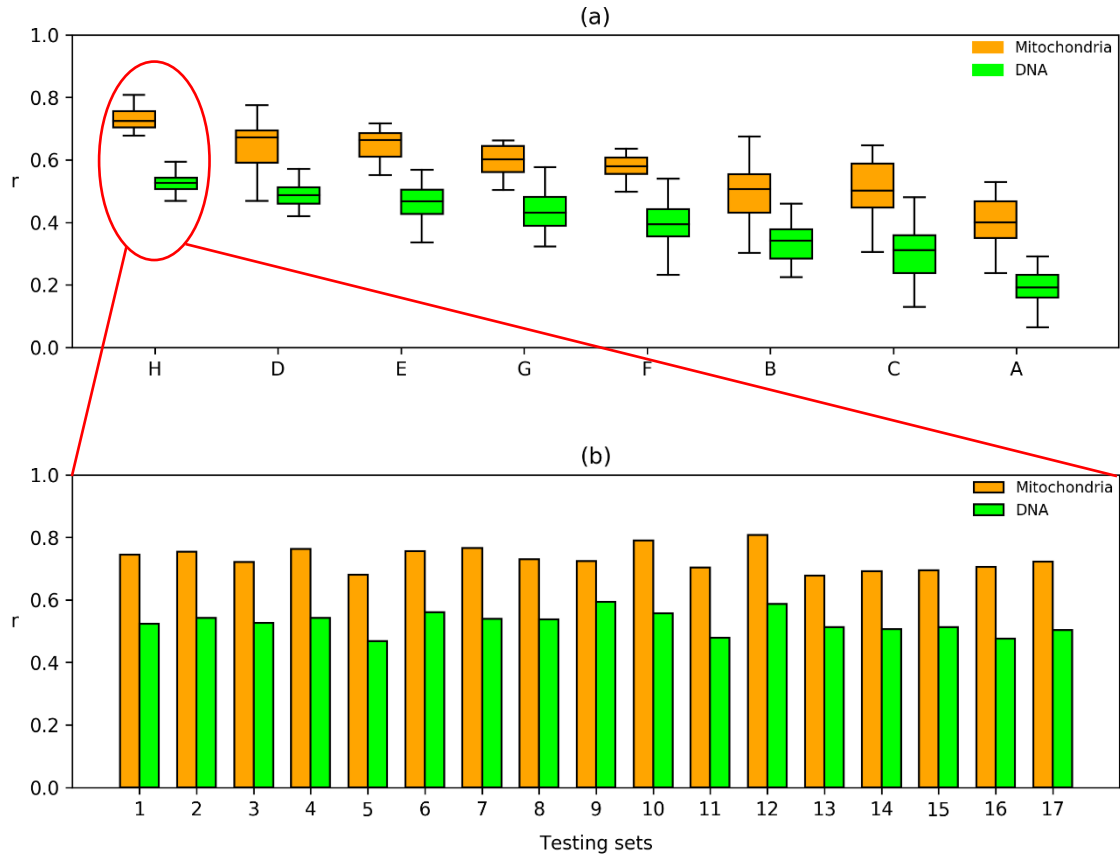


Figure 3.12: Summary results of methods used on this chapter. (a) The performance between different models. (b) The performance from the model with best accuracy (The total number of testing sets are 17). The abbreviation of each model is shown in Table 2.2.

Groups:

A₁ (Mitochondria, Airyscan, 15min)

A₂ (DNA, Airyscan, 15min)

B₁ (Mitochondria, Confocal, 5sec)

B₂ (DNA, Confocal, 5sec)

C₁ (Mitochondria, Airyscan, 1min, HR)

C₂ (DNA, Airyscan, 1min, HR)

D₁ (Mitochondria, Confocal, 30sec, HR)

D₂ (DNA, Confocal, 30sec, HR)

E₁ (Mitochondria, Confocal, z-stack)

E₂ (DNA, Confocal, z-stack)

F₁ (Mitochondria, Airyscan, z-stack, HR)

F₂ (DNA, Airyscan, z-stack, HR)

G₁ (Mitochondria, Airyscan, z-stack, DS)

G₂ (DNA, Airyscan, z-stack, DS)

H₁ (Mitochondria, Confocal, z-stack, HR)

H₂ (DNA, Confocal, z-stack, HR)

HR: high-resolution, DS: downscaling



Section 3-5: Prediction from General Model

In this section, we combined time-series images with z-stack images to train a general (combined) model that can implement in both time-series and z-stack image segmentation tasks. The combined dataset we used to train the model is the same as those we used to train $D_{1,2}$ model and $H_{1,2}$ model. Besides, $D_{1,2}$ and $H_{1,2}$ were also the models we used to evaluate the performance over the general model later on.

Both Fig. 3.13 and 3.14 show that the predictions were successfully conducted in their corresponding models. The difference of the predictions between the z-stack model and the general model was a bit to none, indicating that the general model has almost the same performance as the z-stack model. On the other hand, the difference of the predictions between the time-series model and the general model was more obvious, suggesting that the general model functions worse in the time-series prediction compared to the z-stack prediction.

Fig 3.15 shows the overall performance of each model across different prediction tasks. Note that the predictions from their corresponding models could reach better performances (Combined to combined, time-series to time-series, etc.). In addition, the general (combined) model had relatively fine performance in all three tasks.

Time-series

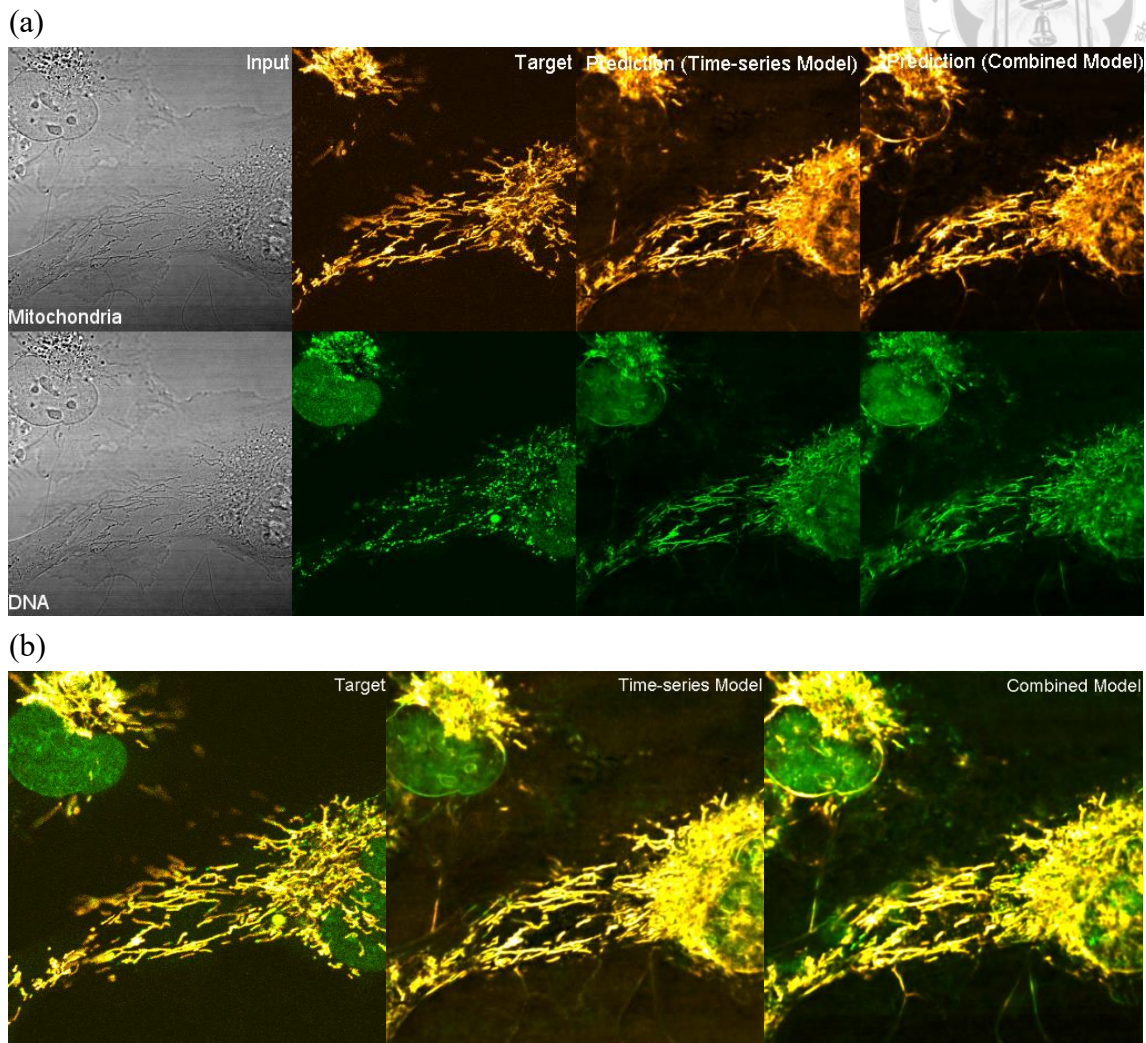


Figure 3.13: (a) Fluorescence prediction from transmitted light across the time-series models and the combined models. The left column is the same single t-slice of a 3D transmitted light input image. The second column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The third column is images predicted by the time-series models. The right column is images predicted by the combined models. The example images are selected from a larger pool of test images. (b) The overlap predictions of each set (Left: Target, Middle: Time-series model, Right: Combined Model).



Z-stack

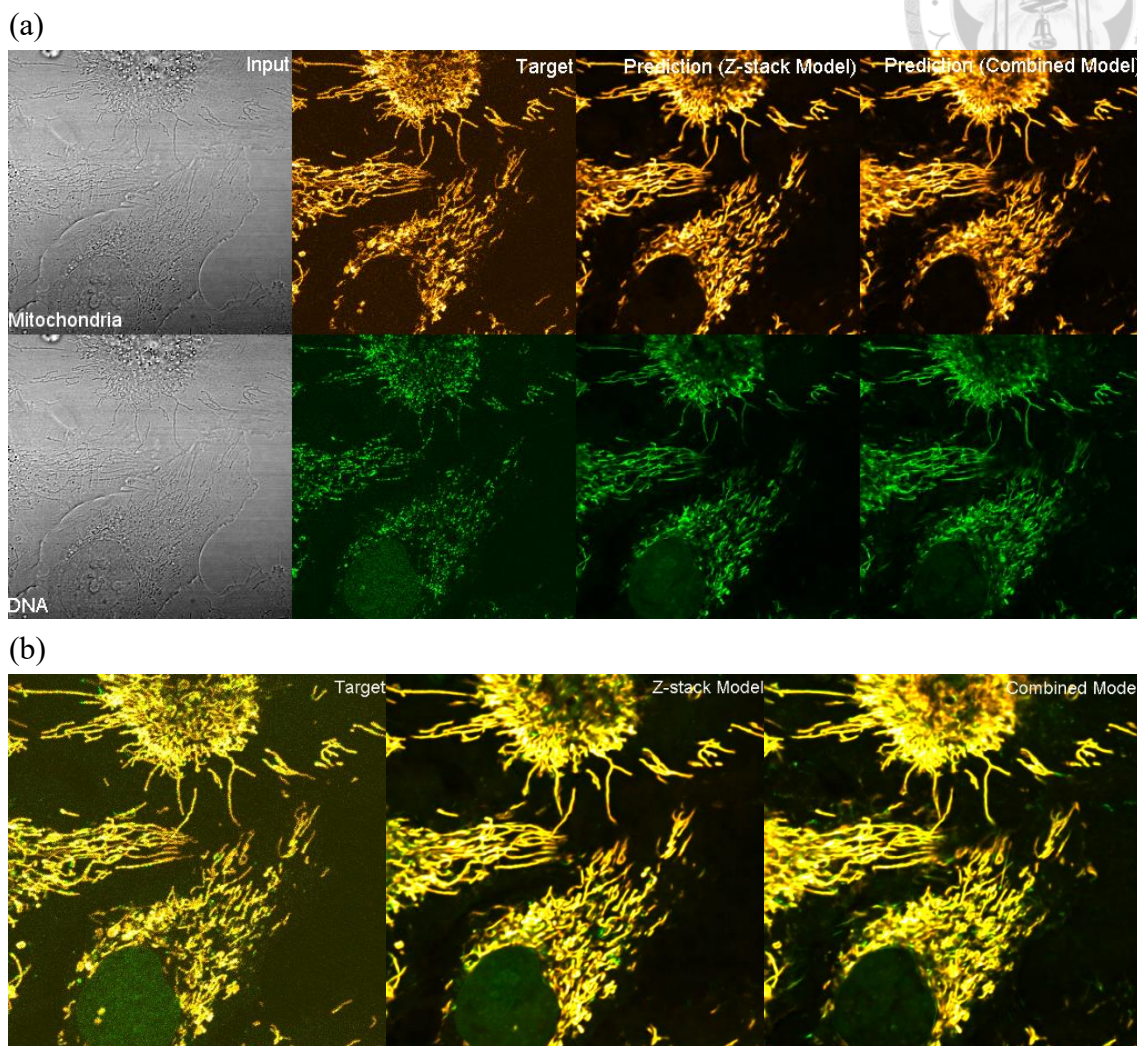


Figure 3.14: (a) Fluorescence prediction from transmitted light across the z-stack models and the combined models. The left column is the same single z-slice of a 3D transmitted light input image. The second column is mitochondria ground-truth (top) and DNA ground-truth (bottom). The third column is images predicted by the z-stack models. The right column is images predicted by the combined models. The example images are selected from a larger pool of test images. (b) The overlap predictions of each set (Left: Target, Middle: Z-stack model, Right: Combined Model).

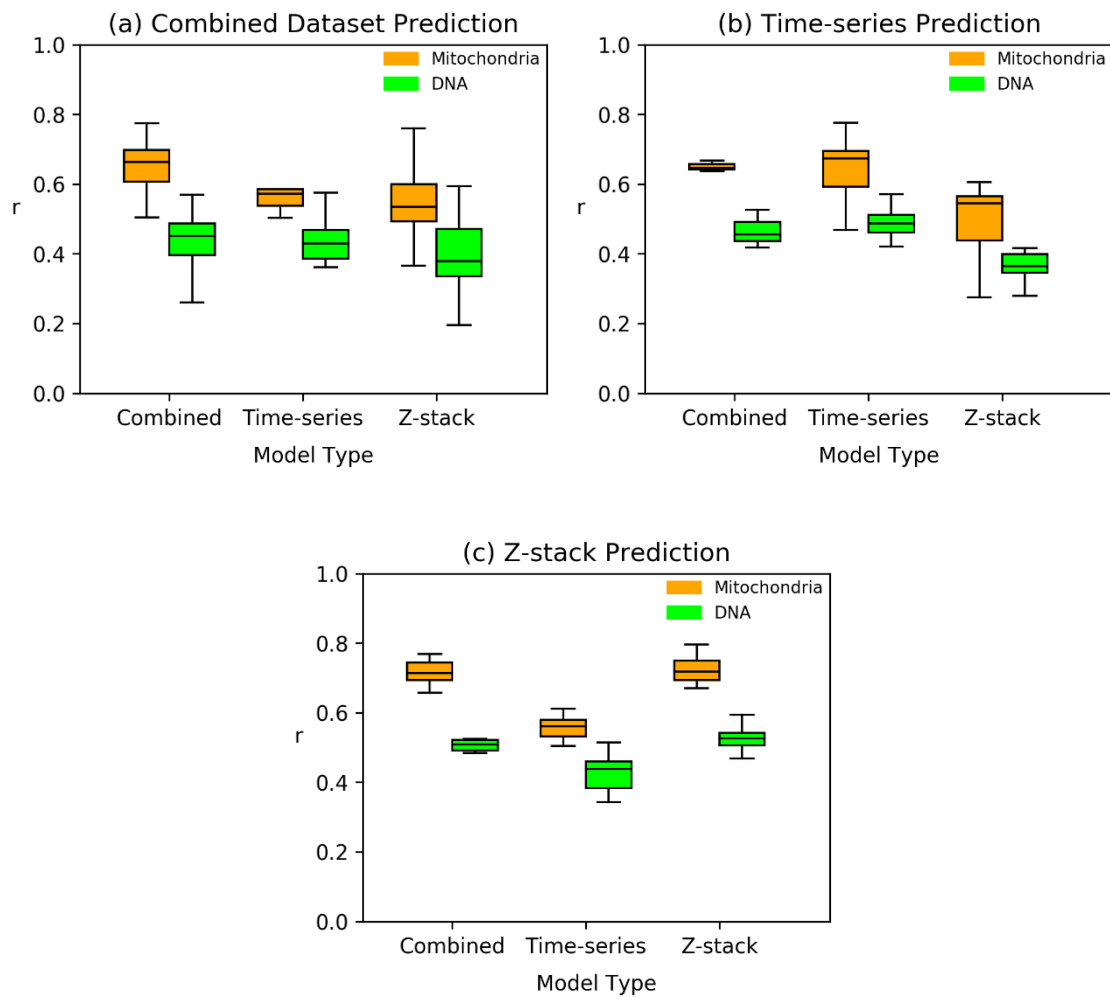



Figure 3.15: (a) Combined dataset prediction performance across different model types. (b) Time-series prediction performance across different model types. (c) Z-stack prediction performance across different model types.

Chapter IV: Discussion



Section 4-1: Label-free Prediction on Mitochondria

The methodology presented in this research provides potential ways to analyze and study mitochondrial structure and dynamics in many aspects. Like previous works done by the Allen Institute for Cell Science[15] and Google[16], it could reduce time sample preparation and cell labeling, eliminating problems of phototoxicity and photobleaching. Moreover, there are two major improvements in our research: (1) Label-free prediction of time-series fluorescence images from TL images. Unlike the time-series prediction from Allen Institute for Cell Science whose training data was static cell images, we used live cell time-series images as our training data. In this way, the correlation between the time-series prediction and actual cell movement will be stronger. As mentioned in Section 1.2, the phototoxicity often happens in long-duration live cell fluorescence imaging, building a barrier for researchers to obtain time-series information. Therefore, our model provides an alternative method to visualize cellular processes with longer timescales. (2) High-resolution prediction from confocal and Airyscan images. Our model was trained with confocal and Airyscan images which provide fine details of mitochondria. Although the prediction performance may drop slightly, such details are valuable for researchers to observe the morphology and dynamics of mitochondria and further analyze them.



Deep learning has been applied in microscope images in recent years, achieving advances in segmentation and classification. For example, using segmented images of simple cells to classify its shape[29], and segmenting cells from bright-field z-stacks[30]. Our research on microscope images using deep learning may contribute to image-based tasks, such as long-time-series observation of tissues or cells[13]. As related works demonstrate the possibility to predict from TL images[15], [16], there is the chance that such deep learning techniques can be applied to other biomedical image tasks.

Section 4-2: Result Analysis

Beside the predictions on Mitochondria and DNA, there were also some factors that affect the model performance. First, to test the effect of number of training images (input data), we trained the model with different numbers of training images, from 2 to 46. The other hyperparameters were remained the same. The dataset used in this experiment was the z-stack confocal images with the resolution of $917 * 917$ pixels, which was the same used in the model H_1 . Fig. 4.1 shows the performance on different numbers of training images. It is obvious that as the number of training images increases, the overall accuracy will improve. However, the performances between 32 and 46 are almost the same, indicating that increasing dataset is not cost-effective to improve the model performance.

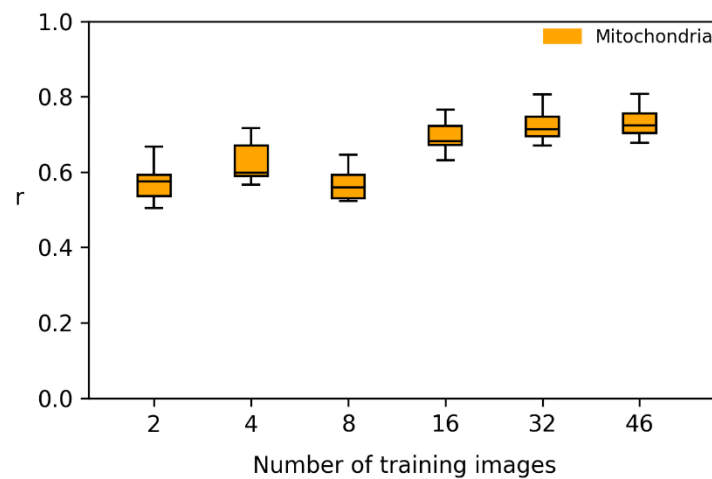


Figure 4.1: Prediction performance across different number of training images.

Second, to test whether the photobleaching will affect the model performance, we evaluated the accuracy between images of different time, from the first to the last. The dataset used in this experiment was the time-series confocal images with the resolution of $917 * 917$ pixels, which was the same used in the model $D_{1,2}$. Fig. 4.2 shows the performance on images of different time. The accuracy in both mitochondria and DNA decreased slightly after long time of imaging. This suggested that the model performance was affected by the photobleaching problem occurred in long-time imaging.

Third, to validate whether the z-stack model can be used in predicting time-series images and thus there is no need for the time-series model, we used two models ($H_{1,2}$ and $D_{1,2}$) which were previously trained with different types of images (z-stack and time-series) to do time-series prediction. Fig. 4.3 shows that the model trained with the time-



series images could achieve better accuracy on the task of predicting time-series images compared to the model trained with the z-stack images. Hence, it is noted that to get better prediction results, the training and the testing images sharing same imaging parameters are recommended.

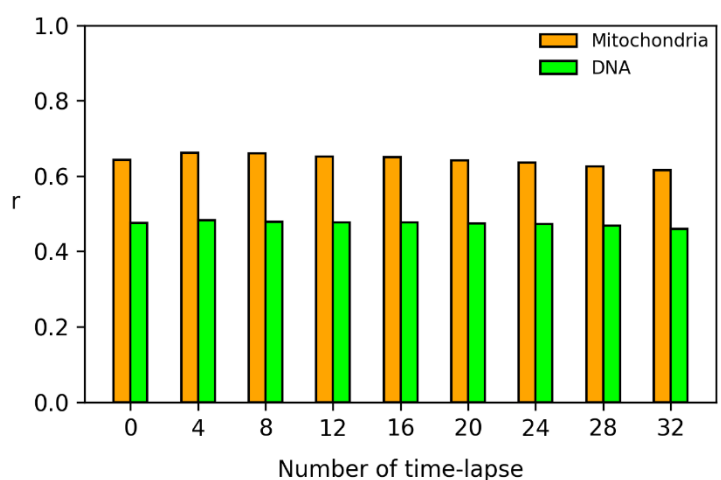


Figure 4.2: Prediction performance across images of different time.

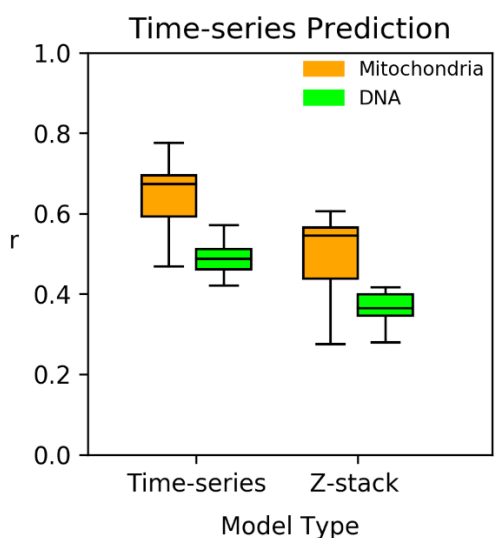


Figure 4.3: Time-series prediction performance across different model types.



Last, the patch size of input data will also affect the prediction. Hence, we trained the model with different patch sizes to evaluate the performance. Here, we used the same model architecture adapted from Allen Institute. The dataset used in this experiment was the z-stack confocal images with the resolution of 917 * 917 pixels, which was the same used in the model H₁. Fig. 4.4 shows that as the patch size increases, the model performance will also improve. Nevertheless, the performance on patch size of 128 was poorer than that on patch size of 64, indicating that increasing the patch size will not lead to better performance unless we modify the architecture of the model.

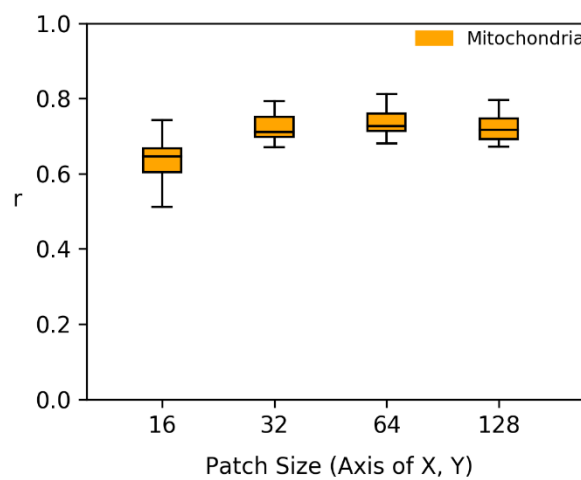



Figure 4.4: Prediction performance across different patch sizes.

Section 4-3: Experimental Difficulties

We faced several difficulties during our experiments. First, the Airyscan microscopy

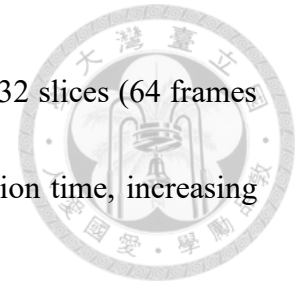


method could not combine with TL microscopy method in the same experimental step. The original system of ZEN Blue 2.6 did not support such acquisition function. To solve this issue, we used “Experiment Designer” inside ZEN Blue 2.6 to create multi-experimental blocks that separate Airyscan and TL microscopy acquisitions into different blocks. In that way, we could acquire TL images and their corresponding Airyscan images simultaneously.

Second, in the experiment of long interval time-series imaging, the cells may move out of the visual field and the color will fade. To deal with the color fading problem, we added 1% antibiotic-antimycotic in each cell medium. But for the cell moving problem, we have not overcome it for yet. The potential solution is to fix the cells with formalin adding into the medium.

Third, for most Airyscan experiments, the acquisition time was so long that kills the cells. To solve the issue, we came up with two possible solutions: (1) Reduce the laser intensity of TMRM channel and SYBR Gold channel. In this way, cells would survive under the acquisition of the Airyscan method. For example, in our time-series experiment of the 15-minute interval, we reduced the intensity of the laser from 1% to 0.2%. (2) Reduce the slices (z-stack) or frames (time-series) of the experiment. In high-resolution

imaging acquisition, we reduced the original 64 slices of z-stack to 32 slices (64 frames of time-series to 32 frames), which can accelerate the total acquisition time, increasing the chance that cells can survive.



Last, in time-series experiments, the water vapor will accumulate on the upper cover of the plate after long-time imaging, resulting in blurs and darkness on TL images. Such noise problem will affect the accuracy of the prediction as we mentioned in Section 3.3. The solution, for now, was to remove the upper cover during the experiment. This move may cause pollution of cells, which needs further improvement and alternative solutions in the future.

Section 4-4: Limitations

The method in this research exists its limitations. First, with supervised machine learning, the prediction performance depends on the information contained in input data. Without clear and correlated TL and fluorescence input, the model cannot learn a relationship between them. For example, in the case of DNA prediction on different conditions, the performance was not good as mitochondria due to the invisibility of mtDNA on TL images and the corresponding low association between TL and fluorescence images of DNA.



Second, the quantity and quality will be key factors that influence our model performance. In most of our experiments, the training data were insufficient since the training process can reach low loss in the training set but unstable loss in the validation set. Besides, in live cell imaging, especially in high-resolution and time-series, there were some dust and water spots on TL images after long time. This will then become the noise on the input data, resulting in poor model performance. The possible solution is to do some denoise procedures before feeding them into the model.

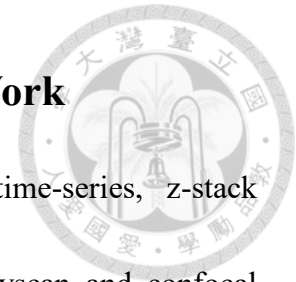
In addition, the model cannot perform as well with inputs of different cellular morphologies and different imaging parameters. In most of our testing sets, the predictions with low Pearson correlation coefficients were those with abnormal shapes. The same issue happened when we used the model from Allen Institute[15] trained with hiPSC to predict our own data (AC16). On the other hand, the prediction from the TL images acquired with imaging parameters that were the same as those used to train the model can achieve the highest accuracy; The prediction from the TL images with different parameters will provide poor results undoubtedly. For instance, the predictions from the TL images acquired with 5-second time interval can achieve ideal performance using the model with 5-second time interval training sets, but achieve bad performance using the

model with 15-minute time interval.



Last, as we know, the model performance can be improved by modifying network architectures and hyperparameters, but it is difficult to understand how such modifications will affect the performance. In our experiment design, there are two networks (Fig. 2.2 and Fig. 2.3) and several hyperparameters need to be decided: patch size, number of iteration, batch size, etc. These all will affect the performance of the model. The only way to find out whether the changes improve the performance is to finish each experiment completely, which spends plenty of time. Hence, having a comprehensive understanding of the functions inside the model will be critical for future improvement.

Chapter V: Conclusion and Future Work

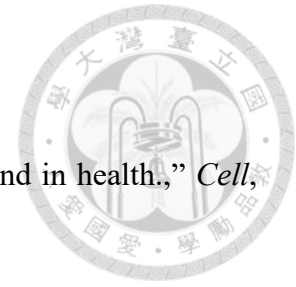


This research constructs specific models that provides time-series, z-stack predictions of Mitochondria structures with TL images from Airyscan and confocal microscopy. In the confocal experiments, prediction results show that both the time-series and z-stack models can achieve relatively fine performance under an interval of short-time or thin z-stack.

To further improve the overall performance, there are some future directions: (1) Increase the total number of TL images and fluorescence images. This can be done by data augmentation such as image rotation or distortion. (2) Data preprocessing. With more stages of preprocessing, such as Salt-and-pepper noise reduction, dust artifact removal, and flat field correction, we can ensure the quality of the training data. (3) Model interpretation. From another point of view, we can try to open the black box of the model, understanding the principle of its prediction work.

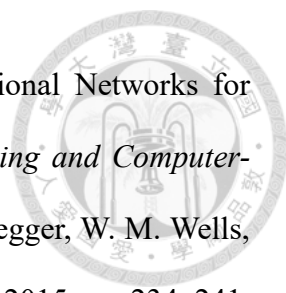
Notwithstanding the limitations existing in our presented methodology, this study does suggest an alternative approach in broader biological imaging areas where it may be an opportunity for a breakthrough.

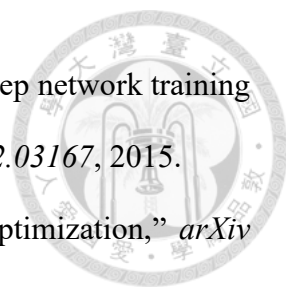
Reference



- [1] J. Nunnari and A. Suomalainen, “Mitochondria: in sickness and in health.,” *Cell*, vol. 148, no. 6, pp. 1145–1159, Mar. 2012.
- [2] D. H. Margineantu and D. M. Hockenbery, “Mitochondrial functions in stem cells.,” *Curr. Opin. Genet. Dev.*, vol. 38, pp. 110–117, Jun. 2016.
- [3] M. R. Duchon, “Roles of mitochondria in health and disease.,” *Diabetes*, vol. 53 Suppl 1, pp. S96–102, Feb. 2004.
- [4] D. C. Chan, “Mitochondria: dynamic organelles in disease, aging, and development.,” *Cell*, vol. 125, no. 7, pp. 1241–1252, Jun. 2006.
- [5] J. Gao, L. Wang, J. Liu, F. Xie, B. Su, and X. Wang, “Abnormalities of mitochondrial dynamics in neurodegenerative diseases.,” *Antioxidants (Basel)*, vol. 6, no. 2, Apr. 2017.
- [6] R. J. Giedt, P. Fumene Feruglio, D. Pathania, K. S. Yang, A. Kilcoyne, C. Vinegoni, T. J. Mitchison, and R. Weissleder, “Computational imaging reveals mitochondrial morphology as a biomarker of cancer phenotype and drug response.,” *Sci. Rep.*, vol. 6, p. 32985, Sep. 2016.
- [7] F. E. Lennon, G. C. Cianci, R. Kanteti, J. J. Riehm, Q. Arif, V. A. Poroyko, E. Lupovitch, W. Vigneswaran, A. Husain, P. Chen, J. K. Liao, M. Sattler, H. L. Kindler, and R. Salgia, “Unique fractal evaluation and therapeutic implications of mitochondrial morphology in malignant mesothelioma.,” *Sci. Rep.*, vol. 6, p. 24578, Apr. 2016.
- [8] D. J. Stephens and V. J. Allan, “Light microscopy techniques for live cell imaging.,” *Science*, vol. 300, no. 5616, pp. 82–86, Apr. 2003.
- [9] V. Jevtic, P. Kindle, and S. V. Avilov, “SYBR Gold dye enables preferential labelling of mitochondrial nucleoids and their time-lapse imaging by structured illumination

- microscopy,” *PLoS One*, vol. 13, no. 9, p. e0203956, Sep. 2018.
- [10] K. Mitra and J. Lippincott-Schwartz, “Analysis of mitochondrial dynamics and functions using imaging approaches,” *Curr. Protoc. Cell Biol.*, vol. Chapter 4, p. Unit 4.25.1–21, Mar. 2010.
- [11] A. Ettinger and T. Wittmann, “Fluorescence live cell imaging,” *Methods Cell Biol.*, vol. 123, pp. 77–94, 2014.
- [12] P. J. Cranfill, B. R. Sell, M. A. Baird, J. R. Allen, Z. Lavagnino, H. M. de Gruiter, G.-J. Kremers, M. W. Davidson, A. Ustione, and D. W. Piston, “Quantitative assessment of fluorescent proteins,” *Nat. Methods*, vol. 13, no. 7, pp. 557–562, May 2016.
- [13] S. Skylaki, O. Hilsenbeck, and T. Schroeder, “Challenges in long-term imaging and quantification of single-cell dynamics,” *Nat. Biotechnol.*, vol. 34, no. 11, pp. 1137–1144, Nov. 2016.
- [14] J. Selinummi, P. Ruusuvuori, I. Podolsky, A. Ozinsky, E. Gold, O. Yli-Harja, A. Aderem, and I. Shmulevich, “Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images,” *PLoS One*, vol. 4, no. 10, p. e7497, Oct. 2009.
- [15] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, “Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy,” *Nat. Methods*, vol. 15, no. 11, pp. 917–920, Sep. 2018.
- [16] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O’Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, “In silico labeling: predicting fluorescent labels in unlabeled images,” *Cell*, vol. 173, no. 3, pp. 792–803.e19, Apr. 2018.

- 
- [17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [18] R. H. Webb, “Confocal optical microscopy,” *Rep. Prog. Phys.*, vol. 59, no. 3, pp. 427–471, Mar. 1996.
- [19] J. Huff, “The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution,” *Nat. Methods*, vol. 12, no. 12, pp. i–ii, Dec. 2015.
- [20] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *Proceedings of the IEEE ...*, 2015.
- [21] J. Dai, Y. Li, K. He, and J. Sun, “R-fcn: Object detection via region-based fully convolutional networks,” *Advances in neural information processing ...*, 2016.
- [22] A. Dosovitskiy and J. T. Springenberg, “Discriminative unsupervised feature learning with convolutional neural networks,” *Advances in neural ...*, 2014.
- [23] B.-C. Chen, W. R. Legant, K. Wang, L. Shao, D. E. Milkie, M. W. Davidson, C. Janetopoulos, X. S. Wu, J. A. Hammer, Z. Liu, B. P. English, Y. Mimori-Kiyosue, D. P. Romero, A. T. Ritter, J. Lippincott-Schwartz, L. Fritz-Laylin, R. D. Mullins, D. M. Mitchell, J. N. Bembenek, A.-C. Reymann, R. Böhme, S. W. Grill, J. T. Wang, G. Seydoux, U. S. Tulu, D. P. Kiehart, and E. Betzig, “Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution.,” *Science*, vol. 346, no. 6208, p. 1257998, Oct. 2014.
- [24] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, “Learning hierarchical features for scene labeling.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

- 
- [25] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.
- [26] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [27] A. Paszke, S. Gross, F. Massa, and A. Lerer, “Pytorch: An imperative style, high-performance deep learning library,” *Advances in neural ...*, 2019.
- [28] S. Dodge and L. Karam, “Understanding how image quality affects deep neural networks,” ... *eighth international conference on quality ...*, 2016.
- [29] Q. Zhong, A. G. Busetto, J. P. Fededa, J. M. Buhmann, and D. W. Gerlich, “Unsupervised modeling of cell morphology dynamics for time-lapse microscopy,” *Nat. Methods*, vol. 9, no. 7, pp. 711–713, May 2012.
- [30] S. K. Sadanandan, P. Ranefall, S. Le Guyader, and C. Wählby, “Automated training of deep convolutional neural networks for cell segmentation.,” *Sci. Rep.*, vol. 7, no. 1, p. 7860, Aug. 2017.