Deep Learning based methods to understand signalling dynamics in intestinal stem cells

MSc Data Science Thesis
Interim Report

Author: Yash Borikar
Supervisor: Dr Deepayan Bhowmik

1 Introduction

Signalling dynamics in intestinal stem cells (ISCs) are essential since they provide valuable insight into fundamental biological processes. ISCs can regenerate themselves, releasing new stem cells from a stem cell pool located in the crypts and developing into specialised cell types including enterocytes, goblet cells, and Paneth cells that comprise the intestinal epithelium. Researchers can uncover critical principles that underlie intestinal stem cells and other types of stem cell biological systems by exploring the complex mechanisms that govern stem cell regulation, cell fate determination, and tissue development [10].

Signalling events that take place for a short timescale from seconds to minutes influences stem cell behaviour, including their determination of fate. The effect of these signalling events on cells takes 5-7 days, including cell differentiation or self-renewal and migration of cells from the crypt base to the luminal surface [7]. Researchers attempting to understand and predict stem cell function present challenges due to this temporal discrepancy.

In vitro and in vivo models are often utilised for stem cell studies. The in vitro, models allow researchers to view and follow biological responses with live imaging capabilities. The in vitro model struggles to adequately replicate the complex and dynamic in vivo environment for a long period of time [6]. Researchers use in vivo models to study cell fate, allowing for a more realistic understanding of stem cell behaviour and fate determination. However, live imaging can be challenging due to factors like tissue complexity, limited accessibility, and technical restrictions, which make it difficult to observe and track signalling events and their long-term effects in real-time [2].

Advancements in deep learning algorithms have revolutionized computer vision, leading to significant breakthroughs in applications like object recognition and tracking [9]. These methods have received a lot of interest in the field of biomedical research because of their ability to explain challenging biological events [4]. One significant application is developing a computer vision tool using a deep learning algorithm for understanding signalling dynamics in intestinal stem cells which can play a vital role in determining the fate of stem cells. This study is important as it may uncover the complex signalling processes that take place within intestinal stem cells by tracking individual cells and tracing their paths over time using the concept of Multiple Object Tracking (MOT). This will help in gaining insights into the cell's movement behaviour and interaction pattern. The algorithm for deep learning will utilize annotated videos of intestinal stem cells along with convolutional neural networks (CNNs) for feature extraction [13]and recurrent neural networks (RNNs) to track the movement of cells over time. It will be trained on a vast dataset of annotated videos to learn the traits that indicate the presence and movement of cells.

2 Aim and Objectives

2.1 Aim

The primary aim of this project is to create highly accurate algorithms based on deep learning, which can effectively detect and analyse cells present in intestinal stem cell videos. The primary focus is tracking these cells' movement and gathering relevant insights for further analysis.

2.2 Objectives

This work will focus on three main objectives as follows:

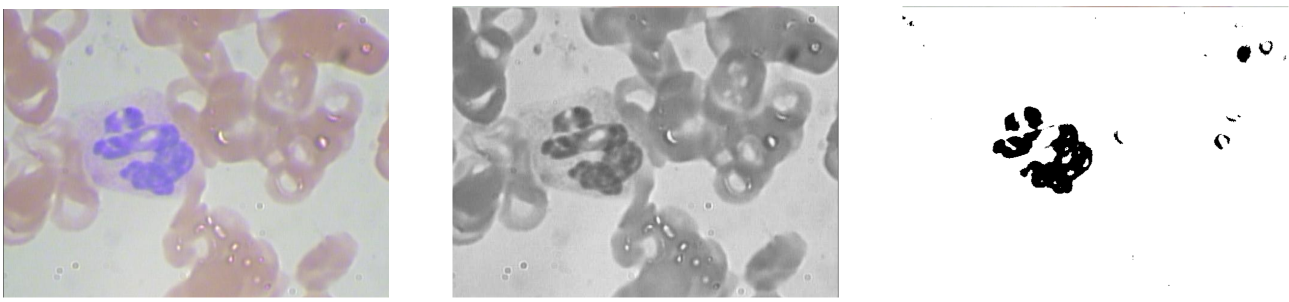
- Primary objective is to develop and train a deep learning model capable of accurately detecting intestinal stem cells in video frames
- Secondary objective is to compare and implement multiple object-tracking models that can precisely trace the motion trajectory of intestinal stem cells [8] [11] [3] [12]. This involves using Kalman filter tracking techniques and the Hungarian algorithm for data association [11] [3] [12].
- Assessing the performance and accuracy of the developed cell detection and path tracing methods.

3 Overview of progress

The in-depth literature review conducted on cell detection and tracing motion trajectory reveals wide recognition of deep learning approaches particularly those using convolutional neural network (CNN) and recurrent neural network (RNN). Fine-tuning pre-trained models or transfer learning on specific datasets, has proven to be effective in achieving accurate cell identification. It is important to acknowledge that these findings mostly apply to general cell recognition and tracking. However, no previous research on intestinal stem cells (ISCs) was not found. This indicates a potential advantage in the application of a deep learning approach in understanding the signalling dynamics of intestinal stem cells towards detection and tracking the path of these cells.

A new and efficient system for object detection framework proposed in [5] is based on Region Proposal Network (RPN), which first generates a set of region proposal and then use the classifier to classify the proposal and regress their bounding boxes. Another popular framework is YOLO (You Only Look Once) [1] which is a single-shot object detector which divides the image into S x S grid cells and predicts the probability of each cell having an object. Paper [4] has proposed a multi-task learning framework that performs both cell detection and cell tracking. [13] introduced cell detection using deep convolutional neural networks to extract features from images and compressed sensing to reconstruct images from features. Simple Online and Realtime Tracking (SORT), introduced in [11] is a new method for multiple object tracking (MOT) which uses the Kalman filter to track objects over time which uses association metrics to link tracked objects in succeeding frame and StrongSORT from [3] is improved performance of DeepSORT that is more robust to motion blur.

Currently, efforts are concentrated on doing extensive preprocessing procedures that involve extracting frames from the video and applying image processing. Image processing includes resizing and normalising the size of each frame and scaling each pixel value between 0 to 1. Adaptive contrast adjustment and unwanted noise reduction using Gaussian or median filtering for prevention of important features. Data augmentation techniques include rotation, translation and flipping to augment the dataset.



(a) (b) (c)

Figure 1: Preprocessing on Blood Cells Image

In the image provided, Figure 1(a) displays the original blood cell image. Figure 1(b) exhibits contrast adjustment, while Figure 1(c) showcases threshold and noise adjustment. These preprocessing techniques can also be applied to Intestinal stem cells for cell identification purposes.

According to research in cell tracking [4], there are several challenges that may arise, such as identifying cell deformation, distinguishing cells from other particles, and dealing with low-resolution images that render cells insignificant in the image.

4 Project Plan

The project timeline is determined from 12th May 2023 to 14th August 2023. Figure 2, illustrates the project plan in detail which includes a preliminary literature review. A partial data preprocessing algorithm is developed on an anonymous cells dataset to effectively detect and track cells. The project's primary implementation phase involves model training, scheduled to begin on 19th May 2023 and end on 17th July 2023. During this interval of time, various model training and evaluation activities will be conducted, following an agile methodology for testing and implementation. The performance of the model will be evaluated on the basis of the accuracy of detecting and tracking Intestinal stem cells.

Project Repository

Git Repository:
`https://github.com/YashBorikar/deep-learning-signalling-dynamics-ISCs`

References

[1] B. Baizhen. Blood cell count and detection method based on yolo. *Highlights in Science, Engineering and Technology*, 27:594–599, 2022.

[2] J. et al. Múnera C. Watson, M. Mahe. An in vivo model of human small intestine using pluripotent stem cells. *Nat Med*, 20:1310–1314, 2014.

[3] S. Yang Z. Yanyun S. Fei G. Tao M. Hongying D. Yunhao, Z. Zhicheng. Strongsort: Make deepsort great again. 2023.

[4] G. Jixiang Y. Zhang H. Tao, M. Hua. Cell tracking using deep neural networks with multi-task learning. *Image and Vision Computing*, 60:142–153, 2017.

[5] R B. Girshick S. Jian R. Shaoqing, H. Kaiming. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[6] M. K. Estes N. F. Shroyer S. E. Blutt K. J. Grande-Allen S. A. Hewes, R. L. Wilson. In vitro models of the small intestine: Engineering challenges and engineering solutions. *Tissue engineering. Part B, Reviews*, 26(4):313–326, 2020.

[7] S. Irshad T. Sandberg D. Worthley S. Leedham S. Biswas, H. Davis. Microenvironmental control of stem cell fate in intestinal homeostasis and disease. *The Journal of pathology*, 237(2):135–145, 2015.

[8] X. Zhou F. Li S. Chen, Y. Xu. Deep learning for multiple object tracking: A survey. *IET Computer Vision*, 13(10):1049, 2019.

[9] Z. Yang L. Wang H. Zhang P. Guo Y. Gao J. Guo S. Guo, S. Wang. A review of deep learning-based visual multi-object tracking algorithms for autonomous driving. *Applied Sciences*, 12(21):10741, 2022.

[10] S. Umar. Intestinal stem cells. *Current gastroenterology reports*, 12(5):340–348, 2010.

[11] P. Dietrich W. Nicolai, B. Alex. Simple online and realtime tracking with a deep association metric. *CoRR*, abs/1703.07402, 2017.

[12] M. Tian R. Yang Y. Qian, H. Shi and Y. Duan. Multiple object tracking for similar, monotonic targets. *2020 10th Institute of Electrical and Electronics Engineers International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 360–363, 2020.

[13] R. Nilanjan Y. Xue. Cell detection with deep convolutional neural network and compressed sensing. 08 2017.

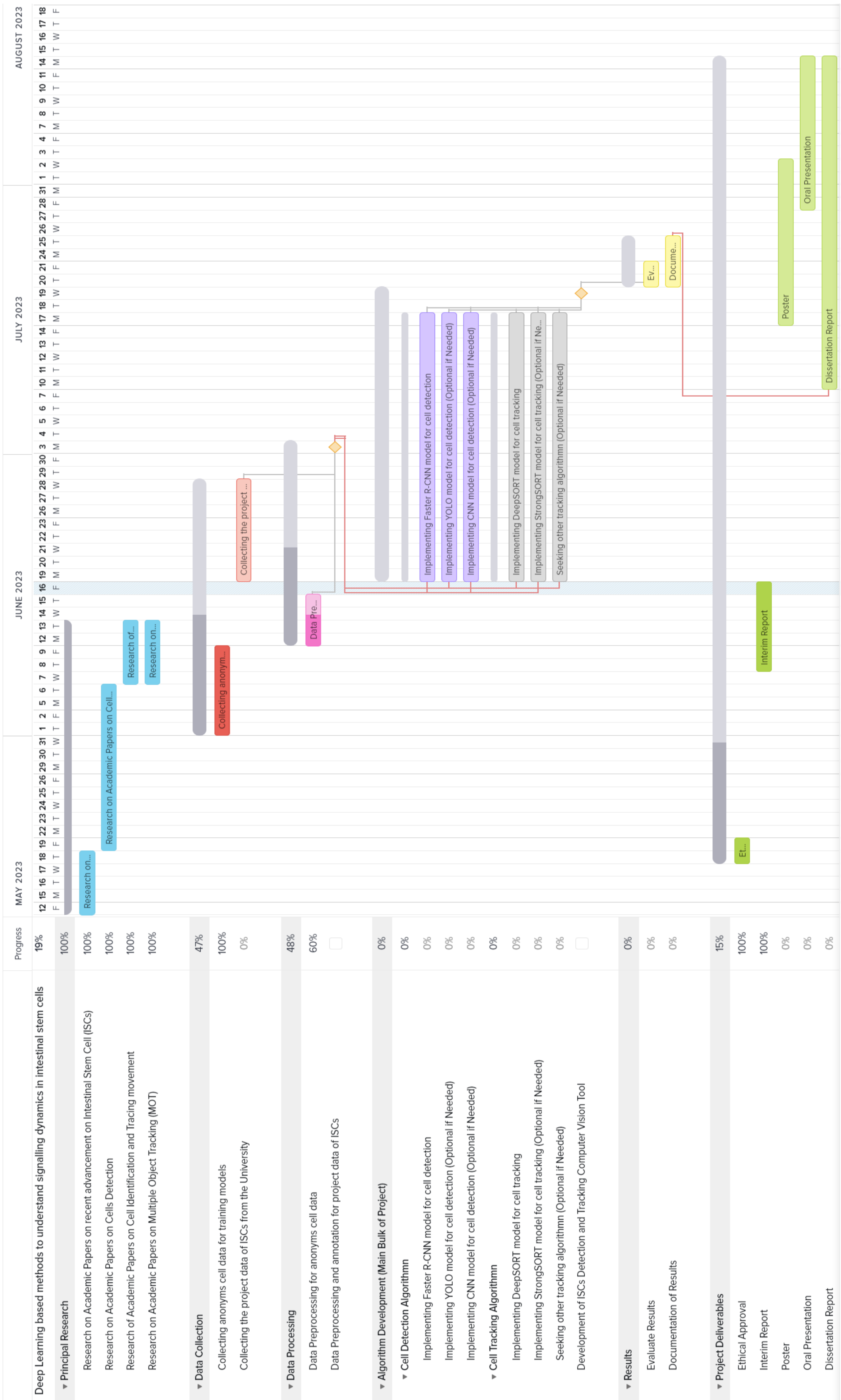


Figure 2: A Gantt Chart of proposed project.

Data Management Plan

0. Proposal name		
Project: Deep Learning based methods to understand signalling dynamics in intestinal stem cells		
Author: Yash Borikar	**Version:** 1	**Date:** 15th June 2023

1. Description of the data

1.1 Type of study

The computational study uses deep learning to develop a computer vision algorithm to detect and trace the movement of intestinal stem cells (ISCs). For this study, a video of intestinal stem cells as a primary data source will be used.

1.2 Types of data

The dataset for this project is now waiting on the University to complete and share it. University's rules for data gathering and collection during this process, including ethical norms and privacy laws.

1.3 Format and scale of the data

Since the complete dataset has not yet been released from the university, a thorough study of it cannot be made at this time. However, progress will be made using open-source data to create algorithms to detect and track anonymous cells. While awaiting the final dataset, this enables approach experimentation and implementation.

2. Data collection/generation

2.1 Methodologies for data collection/generation

This project will acquire its dataset from the University, specifically from the designated data repository or data custodian. Accessing the University's authorized resources, databases, or data sources that contain the essential information necessary for the study will be part of the data-gathering procedure. To ensure compliance with ethical considerations and data protection legislation, the University's data management policies and procedures shall be followed.

2.2 Data Quality and Standards

To maintain the accuracy and reliability of the dataset, a thorough quality assurance procedure will be carried out. This process will involve conscientious checks to detect and rectify any potential inconsistencies or aberrations in the data.

3. Data management, documentation, and curation

3.1 Managing, storing and curating data.

The dataset to be received will be securely backed up on Google Drive and is readily accessible through the Google Colab Pro environment. Additionally, a separate project directory has been set up to store the dataset locally for added convenience.

3.2 Metadata Standards and data documentation

Metadata standards involve capturing, organizing, and describing data. This includes image and video metadata, algorithm details, data collection and preprocessing, cell labelling, training and evaluation schemas, data access and sharing, and long-term preservation will be included in the README file and synchronised with the GitHub repository.

4. Data security and confidentiality of potentially disclosive information

4.1 Main risks to data security

This project's dataset was gathered from the university, assuring its dependability and respect to institutional requirements. The dataset was collected using precise data collection practices, including ethical considerations and quality control techniques. The dataset is collected by certified University researchers and data professionals who adhered to approved techniques and protocols during the data collection process.

5. Data sharing and access

5.1 Suitability for sharing

For this project, careful evaluation of the data collected is required to determine its suitability for sharing by considering crucial factors like privacy, confidentiality, and legal requirements. If the data can be shared without violating any ethical or legal restrictions. Appropriate measures will be made to be easily accessible and facilitate sharing.

5.2 Discovery by potential users of the research data

The suitable metadata will be created and linked to the datasets to ensure the data discoverability. The metadata will include essential details including the data purpose, format, and source. Once the project is completed successfully, an approach for detecting and tracking ISCs will be published. This will aid researchers in their future research endeavours in understanding dynamic signalling in ISCs.

5.3 Data preservation strategy and Standards

Data will be securely stored in a location that is easily accessible and reliable. The retention period of data will be determined, and we will adhere to official preservation guidelines.

5.4 Restrictions or delays to sharing, with planned actions to limit such restrictions.

The README file in the repository will explain the sharing rules and instructions for accessing and using the dataset for research. It will also mention any prerequisites or recommendations for working with the dataset.

6. Responsibilities

Are there any resources required to complete this project?
Required Google Colab Pro+ as we are training the image classification method.
Google Colab Pro +: GBP 45/Month

7. Relevant institutional, departmental or study policies on data sharing and data security

Policy	URL or Reference
Data Management Policy & Procedures	https://www.ncl.ac.uk/media/wwwnclacuk/research/files/ResearchDataManagementPolicy.pdf
Institutional Information Policy	https://services.ncl.ac.uk/itservice/policies/InformationSecurityPolicy-v2_1%20SJ%20v0.1%20amended%202022-08-05.pdf