**PRESENTATION**

Hello everyone. Today I’m here to present my project entitled high-throughput in silico characterization of astrocyte-neuron cell cultures.

This project was conducted under the supervision of Prof. Rui Henriques and is introductory to my Master thesis which will be performed at the Helmholtz Zentrum München at the Götz Lab from the Institute of Stem Cell Research, under the supervision of Dr. Thomas Distler and Dr. Giacomo Masserdotti.

**INTRODUCTION**

The Götz Lab efforts, which I will integrate starting on March, are focused on elucidating the key mechanisms of neurogenesis in the developing and adult brain. In particular, they have pioneered the approach to generate new neurons by direct reprogramming from glial cells which has now become a world-wide very active field of research for novel therapeutic approaches to brain repair.

In this project we draw attention to astrocytes cell cultures.

**MOTIVATION**

These cells are abundant in the central nervous system and are renowned for multiple metabolic, structural, homeostatic functions in the brain, in addition to being a neural cell precursor through a process of dedifferentiation. This way, there arises a compelling need to thoroughly understand their key functions to unravel the mechanisms underpinning neurogenesis.

This necessity has been established by several laboratories, in which the main focuses rely on the detection, segmentation, morphological analysis, spatial profiling, cell-cell contacts, and classification of astrocyte and astrocyte-neuron cultures.

Manual techniques are often hampered by the increasing complexity of the datasets, thus pushing the incorporation of high-throughput automated analysis into modern laboratories. However, automatizing the astrocytes characterization is also currently limited due to their morphological complexity, heterogeneity as a cell population, emphasized by laboratory-specific experimental conditions, and scarcity of annotated data.

**OBJECTIVES AND DELIVERABLES**

The objectives for my Master thesis are divided into three parts. Firstly, I will develop an *in silico* pipeline to characterize astrocyte cultures. Our primary goals will be to perform nuclei quantification, astrocyte identification, and classification of astrocytes based on morphological features and multinucleated state. In addition, we will also potentially aim to characterize cell-cell interactions and perform *in silico* staining of transmitted-light microscopy images.

This project I have conducted in the last semester focused mainly in exploring the fundamental concepts required to perform this characterization and exploring the tools available to do so.

For the Master thesis itself, this work will be pursued and additionally I’ll also start the wet-lab experiments involving the cultivation of murine-derived astrocytes in 3D hydrogels. The final goal will be to establish a high-throughput confocal microscopy platform to streamline the acquisition of images of our cultures and characterize them automatically with our deep learning pipeline, thus accomplishing a high-throughput in silico characterization of astrocyte-neuron cell cultures.

**BACKGROUND**

Machine learning is a branch of artificial intelligence, which deploys several approaches to automatically detect patterns within data. Machine learning is great at learning predictors and descriptors for simple data, such as multivariate data, however, in more complex data domains it usually underperforms. In these scenarios, deep learning is a better approach, since they are representation-based methods that can automatically uncover data representations necessary for each learning task.

So, at the core of deep learning models are artificial neural networks, which are computational models inspired by the structure and functioning of the brain and use perceptrons as their basic units. They possess the ability to effectively model both regression and classification problems by learning network weights that minimize a loss function, associated with the error between the estimation and the targets. Thus, deep learning models are supervised learning approaches.

But, training a robust deep learning network requires a substantial amount of annotated data. In the field of biology this is particularly challenging because data collection is typically labor-intensive and costly. This way, data augmentation, transfer learning and fine-tuning are valuable tools to improve the model’s efficiency.

In the one hand, data augmentation is used to create synthetic data from the original dataset with relatively simple transformations, such as translation, rotation, scaling, cropping, noise injection, etc.

On the other hand, transfer learning is a technique which takes advantage of a pre-trained model to learn a new task faster and with lower amount of data. The traditional transfer learning typically preserves the first layers of a previously trained model and trains the predictor layers with the problem-specific features.

Fine-tuning is a form of transfer learning. Instead of fixing the weight values of the pretrained model, it unfreezes some of the layers phasedly to make them adaptable to the dataset.

In deep learning, there exist diverse architectures designed to respond to specific problem domains. In this project, we will be working with image processing, thus convolutional neural networks are the most suitable architecture.

Convolutional neural networks employ convolution functions to predict an output. Multiple convolutional-polling layers allow to discern complex feature relations and predict an output.

This way, CNNs are characterized by their sparse interactions, parameter sharing and equivariance to translation.

**RELATED WORK**

In this project I’ve conducted a thorough literature review over the most used CNN architectures for image classification, object detection, image segmentation and *in silico* staining.

In **image classification** tasks, the goal is to assign a target label to each observation. VGG and ResNets were singled as the two architectures most used for image classification. Nevertheless, they have also been successfully adapted for other computer vision tasks such as object detection and image segmentation.

On the other hand, **object detection** is the task of pinpointing objects within an image using bounding boxes and assigning corresponding labels. R-CNNs and YOLO networks are the most popular approaches due to their stable performance and coherent results across multiple validated datasets.

R-CNNs are region proposal algorithms combined with CNNs. This approach involves three distinct modules. The first module generates region proposals, and it is followed by a large CNN to extract feature vectors. The last module does object classification using class-specific linear SVMs.

Faster R-CNN, is a later improvement of R-CNN and consists of two networks that can be optimized end-to-end with higher efficiency.

YOLO on the other hand, frames object detection as a regression problem. It looks at the whole image at the same time allowing real-time object detection. There are already available new versions of the originally proposed. In particular, I found that YOLOv5 and YOLOv8 have been used in the literature for tasks similar to the ones I aim to employ.

Moving to **image segmentation**, its purpose is to perform pixel-wise classification, discerning objects from the background. It can be casted into 2 different tasks.

Semantic segmentation simply aims to distinguish background from objects categories. Deep lab and SegNet, are based on ResNets and VGG architectures and perform this task. However, in this work we are mostly concerned with instance segmentation task as to delineate each instance of the image separately.

For this, U-Nets are acknowledged to be one of the best approaches. The traditional U-net has a two-dimensional fully convolutional architecture, consisting of an encoder, followed by a decoder. It is very efficient at training due to few training parameters and yields very precise segmentations.

It served as a template for multiple segmentation deep networks, including StarDist and Cellpose, which I have employed in this work.

Stardist works by approximating the bounding box surrounding each cell instance to a star-convex polygon. Cellpose on the other hand generates topological maps to segment cell nuclei or cytoplasm instances. Additionally, Mask R-CNN takes a different approach from the previous and it starts by determining the objects’ bounding boxes by employing the Faster R-CNN algorithm, and then distinguishes the background from the foreground.

Finally, **in silico staining** represents a pixel-wise regression task that aims to predict fluorescent markers. Bright2Nuc is a network developed by Carsten Marr Lab from the Helmholtz Munich that was designed to predict cell nuclei staining of 3D stem cell cultures from z-stack brightfield images acquired with confocal microscopy.

This network was built upon InstantDL, which is pipeline developed by the same group and empowered with classification, segmentation and pixel-wise regression algorithms.

**SOLUTION PROPOSAL**

As already stated, the final objective of this project is to build an *in silico* characterization pipeline for astrocyte cultures, with potential to characterize astrocyte-neuron co-cultures.

To start off, I will just focus on the nuclei quantification and astrocyte classification tasks, leaving cell-cell interactions and in silico staining approaches for later.

The dataset kindly provided by Dr. Thomas Distler and Dr. Giacomo Masserdotti, consisted of 21 3D z-stacks of confocal microscopy of murine-derived astrocyte cultures embedded in 3D hydrogels and marked for GFAP and DAPI.

So, for the nuclei quantification my proposal is to segment each nuclei instance thus obtaining the total number of cells in the culture. Is important to note that the final objective is to perform 3D characterization, however, in this project I have attempted to do a 2d characterization due to its simplicity and to gain a better insight of our dataset and the feasibility of the proposal.

This way, I’ve designed an automatized preprocessing pipeline using fiji python package and generated 4 auxiliary datasets, both of 2d projections and 3d stacks, containing the DAPI or DAPI and GFAP markers.

For the 2D nuclei segmentation approach, I’ve used the 2D nuclei dataset that I have manually annotated and InstantDL pipeline, specifically Mask R-CNN algorithm with several fine-tuned models.

The results were validated using Cellpose 2d and stardist 2d algorithms, and these metrics were used to evaluate the models’ performance.

For the 3D nuclei quantification, I plan on using the stardist 3d algorithm and validating the results using u-net 3d and cellpose 3d.

As for the astrocyte classification tasks, I’m dividing the problem into three stages attemptively always using 3D images. I aim to harness YOLOv8 and StarDist 3D for these tasks. Ultimately, the models’ efficiency will be assessed using the accuracy, the mean AP and the mean IoU as suggested by previous works.

**CONCLUSION**

In this project I’ve laid a structured problem definition and solid groundwork for the forthcoming efforts. In addition, I have also accomplished to elaborate a detailed solution proposal. My future endeavors will focus on pursuing the elaboration of the characterization pipeline and start the laboratory work with the cultivation of astrocyte cultures.