Project 12 : Artificial Intelligence to quantify the neurodegeneration

Project owner:
BROCARD Jacques (ENS)

Tutor:
MEYER Alexandre (LIRIS)

Students:
ALIZADEH-NOIRET Grégoire
DESAINTJEAN William
LAHAIE Thomas
LEON FOUN LIN Ravy

Master 2 BioInformatique - 2022/2023

Table of Contents

1 Project Presentation

1.1 Actor Participants

1.2 Context

Lipid dysregulation research is a new approach to understand mechanisms behind neurodegenerative diseases such as Parkinson's or Alzheimer's diseases. To study Age-Related Macular Degeneration (ARMD), another neurodegenerative disease that consists of the loss of central vision which appears with age, a LBMC researchers team led by Bertrand MOLLEREAU use *Drosophilia melanogaster* (fruit fly) as a model for this work.

The experimental strategy is to knock out some gene of interest (e.g Fatty acid transport protein 1 or FATP1), then to measure the phenotypic impact of these genes on eye tissues. Precisely, they focus the measurements on the ommatidia (fig 1), the functional unity of arthropods eyes.

(figure ommatidia/microscope)

Drosophilia sp. have average 800 ommatidia per eye, each ommatidium is composed by a photoreceptor cell cluster of eight cells (enumerate from r1 to r8) :

- r1 to r6 : photoreceptor cell that expresses Rhodopsin 1 (RH1), for scotopic vision.
- r7 and r8 : other photoreceptor cells.

If the inactivation of a gene of interest leads to a neurodegeneration, photoreceptor cells will die and ommatidia development will be affected.

To check the survival of these cells, they performed a confocal fluorescence microscopy. This is made possible thanks to the gene-fusion with Green Fluorescent Protein (GFP) which is used as a reporter of RH1 expression (GFP is placed under the control of the RH1 promoter). Thus, if RH1 is expressed, the photoreceptor cell is alive. (ref technique voir sujet 12).

The phenotypic characterization is based on the number of photoreceptor cells per ommatidium. This quantitative approach allows the use of statistical tools and provides robust results. However, the cell counting is currently hand-made, laborious and experimenter-dependent.

In order to develop an automatic counting software of photoreceptor from microscopic image, Jacques BROCARD is looking for a convolutional neural network, to identify and count cells.

1.3 Objectives

The aim of this project is to use a Machine Learning approach to count the number of photoreceptors on a picture, at the .tiff format, of a wild type drosophila ommatidia. We will use images made by the experimental team, which was annotated by an image analysis program, to train our models. Once the training on the wild type and mutant drosophila ommatidia pictures is complete, and has good acceptability criteria, we will use this model on a test set of pictures. Finally we must evaluate our models by computing the loss function and the confusion matrix.

1.4 Ressources

For the project, numerous pictures of drosophila eyes have been taken by the team of Dr. Bertrand MOLLEREAU by using confocal microscopy: for one eye, several pictures are made at different depths (z axis) to see every receptor of the retina that has the form of a dome.

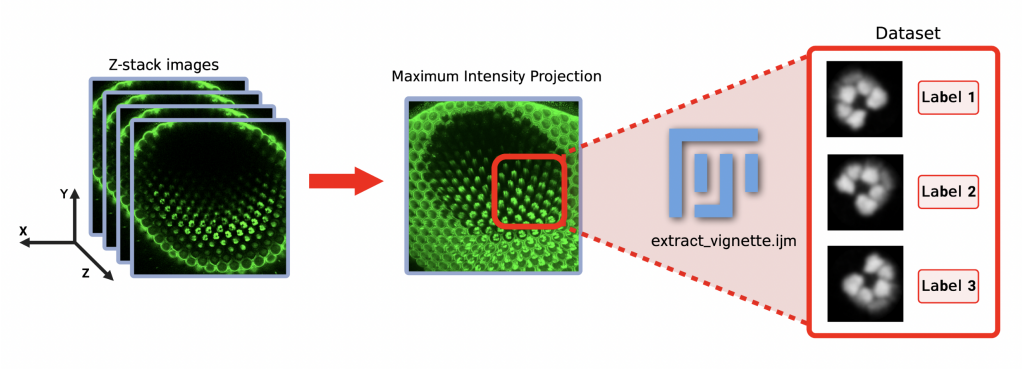

Figure 1: Images made by the experimental team. The first images represent the stack of images taken at different depths (the Z-axes). The 2nd image is a reconstruction of the original image using the stack of images. And the last 3 pictures, represents the vignettes and their labels, made with ImageJ from the reconstructed images.

After that, Jacques BROCARD used Fiji ImageJ [1], an open source image processing package, to superimpose all the pictures on the Z axis to see all the ommatidia of the eyes in one image. Then the software asks the user to place each ommatidium in a circle of adapted size, and produces a vignette of 60 x 60 pixels. These vignettes, all of the same size, are the images analyzed with machine learning.

The expert was not available to label the images at this time of the project. Jacques BROCARD used Fiji ImageJ again to compute the number of receptors, represented by several whites pixels in a circular form on the images. For that, he counts the number of stains by decreasing the gradient of color on the image's pixels, from white to black. This program makes a lot of errors, but it's a first labeling to work on deep learning. The idea is that experts will label the images when they will be free, after the project.

A google collab already exists, using Tensorflow with the library Keras. But there are other wild number of python libraries to use machine learning techniques (e.g. sklearn).

Jacques BROCARD gave us a notebook containing a basic functional neural network. We could try adapting this notebook to work with our data as a first step.

1.5 Acceptibility Criteria

No criteria were explicitly specified by the researchers, but it is known in Artificial Intelligence (AI) that a good model has a certain balance between bias and likelihood. To evaluate that we will focus our work on having a model with the best loss function, and a balance between specificity and sensibility. This will be done by computing the confusion matrix and the ROC curve.

2 Functional needing

2.1 Functional Requirements

Requirements	Must	May
Count the number of photoreceptors in an ommatidia	X	
Detect ommatidia in an image and crop them		X
Remove noise from an ommatidium image		X

2.2 Non Functional Requirements

Requirements	Must	May
The network's predictions must have a correct accuracy (70 90 percent)	X	
User-friendly	X	

3 Constraints

3.1 Deadlines

3 October to 14 October: first period of project

14 October: deadline for specifications

31 October to 11 November: second period of project

28 November to 9 December: third period of project

The final version must be delivered on the 15th of December.

3.2 Technical Constraints

We do not have numerous data and they have incorrect labels. We will make the model with them anyway, but the model will need to be trained again with the right annotations made by the experts, the persons from the experimental part. The time allowed for the project is also short.

We are not limited by the power of our machine: the google collab notebook allows us to launch our different trials of machine learning on google's servers, even with GPU, for free. This will be sufficient to use the different methods, even the most resource-consuming.

4 Project phases

4.1 Planning

03/10 - 14/10 : First meeting with the researchers and the tutors, start of the making of the models to see the different possibilities, and writing of the bill of specifications.

30/09 - 11/11 : Creations of the models: each member of the group will focus his work on the implementation of a method. If we can make them work with our datas, we must solve the problems that can occur, like the too low number of labeled images for the learning: we must evaluate the necessity of increasing data.

28/11 - 09/12 : Evaluating and comparing the models. Redaction of the final work and preparation for the defense's project

4.2 Quality assurance plan

To assess the quality of a given network, we will need to check for several metrics: Its sensitivity (detected positives / all real positives) and specificity (detected negatives / all real negatives) The Receiver Operating Characteristic (ROC) curve, which displays the relationship between the true positive rate and the false positive rate

4.3 Final Product

The final product will be a python Notebook. It will take for input a set of images in 60x60 of an ommatidia made by electronic microscopy and will return the number of photoreceptors with a probability of confidence. The code will be available on GitLab with a README. We must add to the GitLab a written report to explain our choices to implement the differents methods, and to compare the results with differents indicators.

Bibliography

[1] Johannes Schindelin et al. 'Fiji: an open-source platform for biological-image analysis'. In: *Nature methods* 9.7 (2012), pp. 676–682.

BIBtex in progress...