MULTRISPECTRAL MICROSCOPY IMAGE SEGMENTATION WITH U-NET FOR CHARACTERIZATION OF L-FORM BACTERIA

Luc GABEL
Ingénierie des Sciences du Vivant

Théo Nass
Ingénierie des Sciences du Vivant

Project and data provided by : Laboratory of Microbiology and Microtechnology, EPFL

Abstract

The aim of this project is to get quantitative data on L-form bacteria (uropathogenic *E. coli* treated with fosfomycin in Synthetic Human Urine medium) such as their position, length, width etc. from multi-channel microscopy images, in order to quantify the L-form bacteria formation, killing and regrowth over time during antibiotic treatment and subsequent washout phases.

We present a pipeline for a supervised learning approach using a fully connected convolutional neural network (U-net) architecture for semantic segmentation, followed by object classification and quantification using standard image analysis techniques.

1 Introduction

Urinary tract infections (UTIs) are among the most common bacterial infections, affecting about 150 million people worldwide each year and generating substantial healthcare costs. The vast majority of community-acquired infections are caused by uropathogenic strains of Escherichia coli, termed in short UPEC. The gold standard treatment course remains to date the administration of antibiotics, the first-line drugs of choice being trimethoprim-sulfamethoxazole, nitrofurantoin and fosfomycin. The latter is a cell-wall inhibitor, which acts by blocking the synthesis of the bacterial cell wall. This mesh-like structure is made of peptidoglycan and is located — in the case of Gram-negative bacteria such as E. coli — in between the inner and outer plasma membranes. It gives bacteria their shape and helps them endure the turgor pressure caused by most growth media having a lower osmolarity than the bacterial cytoplasm. This causes bacteria to lyse when treated with a cell-wall inhibitor in conventional growth media. However, in a hypertonic medium such as human urine, bacteria can survive the loss of peptidoglycan caused by the action of the antibiotic and turn into so-called 'L-form bacteria', where the cell wall structure is gone and the inner and outer membrane are free from one another. As these membranes are more fragile than peptidoglycan, most L-form bacteria eventually also die or at least cannot regrow when the antibiotic is removed, yet a small fraction of them manage to revert back to their original shape and proliferate, leading to a resurgence of the disease.

In order to combat this resurgence, it is important to understand what differentiates bacteria that are able to revert back and proliferate from those that are not. To get data on this phenomenon, uropathogenic E. coli expressing GFP (green fluorescent protein) were grown in Synthetic Human Urine medium in a custom-made microfluidic chip, into which fosfomycin was flown and then washed out. The formation of E. coli L-forms was observed, and the chip was imaged on phase-contrast and green fluorescence channels every 5 minutes for 12 hours in total, at which point some E. coli had reverted back to their original shape and started proliferating, leading to the formation of bacterial microcolonies inside the device.

2 Method

2.1 Data Exploration

The original time-lapses were in .ndi format (Nikon microscope format), each containing 144 .tif images with 3 channels: Phase (bright-field), GFP (fluorescence linked to Green fluorescent protein) and TRITC (a dye used to visualise the addition and wash of the antibiotic).

Using ImageJ (a common bio-analysis software), we selected only the Phase and GFP channels. We took the frames 34 (antibiotic has just been washed out) to 63 (beginning of regrowth of reverted bacteria) of each time-lapse, that is to say we analyse the 2h30 of time-lapse where the bacteria are all in L-form shape. These images were renamed with their time-lapse number and a new frame number going from 000 to 029 instead of 34-63 , but the order was

conserved. They were saved as separate .tif files.

This newly created dataset was first analyzed using a bio-microscopy image analysis software called Ilastik. This software uses a series of features (filters) to perform semantic and/or instance segmentation of cells with only a few hand drawn annotations. As seen in Figure 1, this software gives a relatively good semantic segmentation for qualitative analysis, but is not precise enough for single cell quantitative measurements as it does not allow for instance segmentation.

To increase the quality of our segmentation, we decided to use a supervised learning approach. Hence we had to annotate our images. Seeing as free software for instance segmentation were not readily available, we chose to do a semantic annotation, and use image analysis methods for object classification on the segmentation masks created by the supervised learning approach.

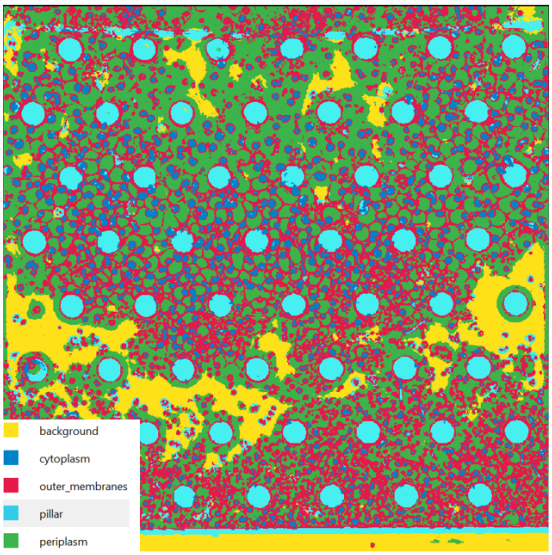

Figure 1: Segmentation map using Ilastik

2.2 Ground truth annotation

Phase images were hand-annotated (semantic segmentation) using the raster graphics software Krita. The cell ground-truths were painted in a transparent layer on top of the phase image. In the resulting mask, the objects were determined by the alpha-channel value (from 0, i.e. transparent, for background, to 1, i.e. opaque, for cells).
The annotation process is extremely time consuming, annotating a single image takes 1h30-3h in order to circle the 500-1000 cells present in each image. This is why we annotated only 6 images. To still capture the heterogeneity of the bacteria's shape during the experiment (the L-form bacteria grow) as well as the variation in lighting or other factors between time-lapse, we chose to annotate the first and last images of 3 time-lapses.
The ground truth annotation for the GFP images was done using Ilastik. This was possible because of the high contrast of the GFP images.

2.3 Preprocessing

2.3.1 Registration

In order to modify all the images at once, they needed to be registered. We cropped the upper left corner of all images to get an image with a single pillar. This allowed us to use pattern matching between a circle and the pillar's halo to get the position of the center of the pillar. The shift in between frames was calculated by subtracting the position of the pillar in each frame to that of the first frame. The images were the registered according to their shift.

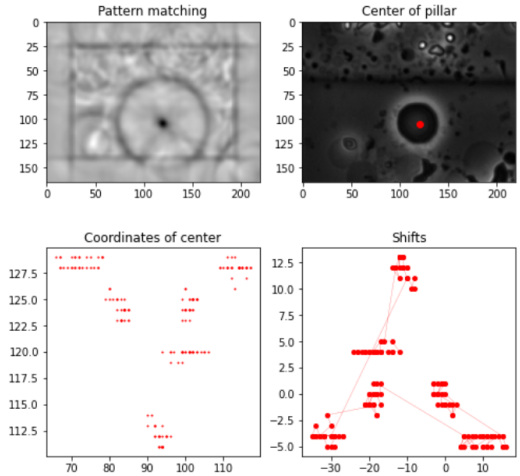

Figure 2: Graph or center position, pattern matching and shifts

2.3.2 Gaussian blur

As you can see on the phase images, pillars are enveloped by a bright halo, which makes the detection of cell close to the pillars harder. To decrease the brightness of this halo, we performed a Gaussian blur (sigma=7) on the phase images. The blurred images were then subtracted from the original images. This operation basically equates to applying a low-pass filter on the phase image.

2.3.3 Binary mask

The transparent mask was converted to a single-channel binary mask (0 for background, 1 for cells).

2.3.4 Splitting

In order to avoid excessive resizing which would lead to loss of information, and to bring the images to a format more similar to known successful examples, each image was split into 4 sub-images.

3 Supervised Learning with U-Net

3.0.1 Data Augmentation

5 operations were implemented to augment the dataset : rotations at 90°, 180° and 270°, along vertical and horizontal flipping. Because of the pillars, we decided not to implement deformations in the images.

3.0.2 U-Net

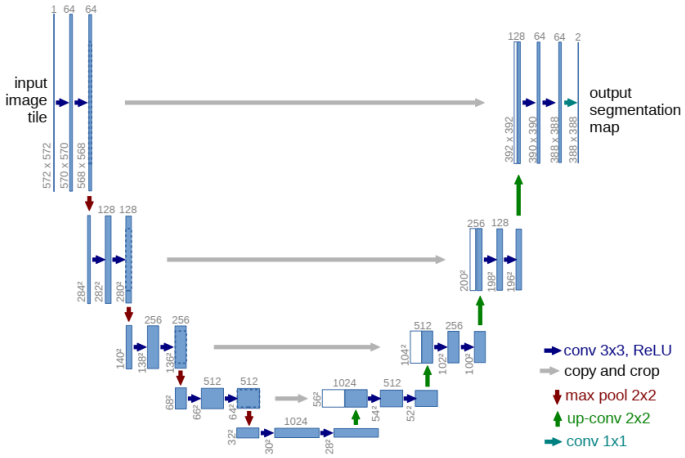

Figure 3: U-net Architecture, Ronneberger et al, 2015
U-net is a fully-convolutional neural network created by Ronneberger et al in 2015 [3]. Its name comes from its U-shaped architecture. It is composed of a contracting path (left side) and an expansive path (right side). The contracting path consists of consecutive 3x3 convolutions, each followed by a ReLU activation step and a 2x2 maxpooling step. The expansive path consists of 2x2 upconvolutions, concatenation with the corresponding feature map from the contracting path, then 3x3 convolutions.

This network has shown good results on cell segmentation, even with small datasets. This network won the ISBI cell tracking challenge in 2015 [3]

Adam was chosen as the optimizer for its fast convergence and computational efficiency.

3.0.3 Choice of Loss Function

Several loss functions were considered and implemented for this project.

- Binary crossentropy
- Dice loss
- Focal loss

Binary crossentropy is a common choice of loss function for image segmentation applications. It was also the recommended loss function in the original U-Net paper [3].

Dice loss is based on the Sørensen–Dice coefficient, also known as the F1 score. The Sørensen–Dice coefficient is a statistic used to estimate the similarities of two samples.

$$\text{SDC} = \frac{|\text{True} \cap \text{Predicted}|}{|\text{True}| + |\text{Predictet}|}$$

The associated loss function is of the form :

$$\text{Dice loss} = 1 - \text{SDC}$$

The Dice loss function was created for segmentation problems where the classes were strongly unbalanced[2], which is the case here.

Focal loss is a variant of the crossentropy loss function [1]. It was introduced to handle class imbalance in the object detector RetinaNet. Focal loss assigns more weights to examples that are hard to classify,which is more suited to cases where the class imbalance is strong. It contains a hyperparameter γ which can be otpimized for the task at hand.

All three loss functions were implemented and tested in our model.

4 Object classification and Quantification

The U-net model creates a binary segmentation map. However, we are interested in instance segmentation instead of a semantic one. To bridge the gap, we transform every segmented cell (pixel of high intensity surrounded by pixel of low intensity) into an object, with its own label. The newly created image represents every separate object with the intensity equal to the label (background = 0, first cell = 1 and so on). We then extract the following characteristic for every cell: label number, position (x and y axis), length, width, perimeter and area.
To turn these features into actionable data, we need to make a few modifications. The shift, which was induced by the registration step (see Registration) is subtracted to get the position of the center in the non-registered image. Furthermore, all lengths and surfaces, which are in pixels, are converted into μm using a pixel to μm ratio of 4.633. This ratio was obtained using ImageJ's Set Scale functionality with a distance between the center of the pillars of 30μm. The final output is a .csv file. It is named after the time-lapse and frame number, and contains all the characteristics mentioned above for each cell.

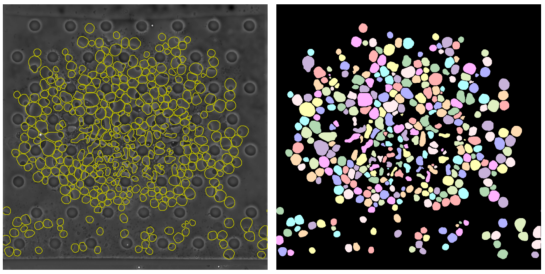

Figure 4: Example of instance segmentation from semantic segmentation mask

5 Result

5.0.1 Training : Short-term Behaviour

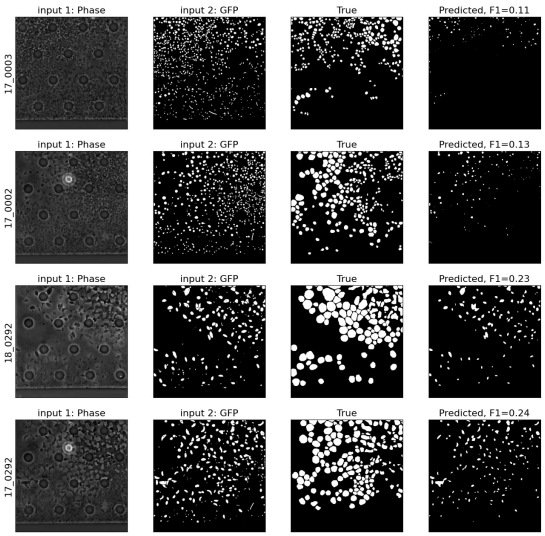

Figure 5: Phase (column 1), GFP (column 2), Ground Truth (column 3) and Predictions (column 4) after 10 epochs, with Dice loss
Short training sessions (< 20 epochs) showed that the CNN was learning primarily on GFP data.

5.0.2 Training : Long-term Behaviour

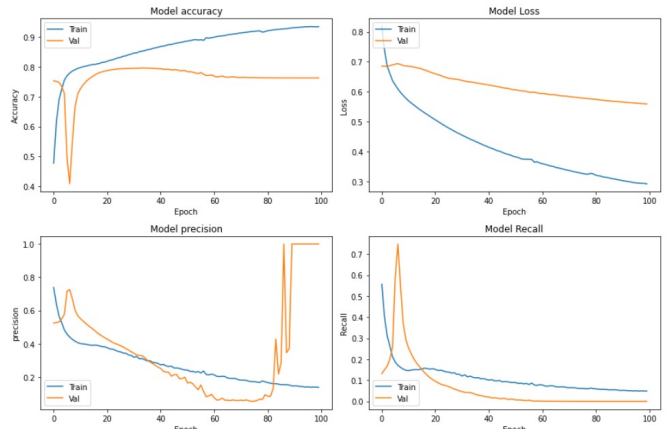

Figure 6: Model metrics
The precision and recall systematically decay over time, eventually outputing an empty mask. This problem occurred with every loss function, independently of whether the data was augmented or split.

5.0.3 Potential improvements

To improve on this, the first step would be to generate more annotated data.
Various parameters may also be optimized :

- Batch size for Stochastic Gradient Descent
- The hyperparameter $\gamma$of the focal loss function
- Removing artefacts in the microscopy images

6 Conclusion

Our project is a first step to a functional segmentation network. The data preprocessing was carefully crafted to address the unique characteristics of our dataset. The object classification and quantification has been shown to work perfectly using ground truth images. Given the impressive track record of U-net in other biological segmentation challenges, we believe that, with sufficient optimization and troubleshooting, and a bigger annotated dataset, our project could become an effective tool in characterization of L-form E.coli. It would still be interested to consider more basic approach for segmentation, such as K-means on the phase images or spectral clustering on a composite image of GFP and phase.

References

[1] T.-Y. L. et al. Focal loss for dense object detection. *Facebook AI Research*, 2017.

[2] F. Milletari, N. Navab, and S. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. pages 565–571, 2016. doi: 10.1109/3DV.2016.79.

[3] P. F. Olaf Ronneberger and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.