

A Deep Learning Pipeline for Genome-Wide Imaging Screen Uncovering Cell Death Regulators

1st Salma Kazemi Rashed

*dept. Experimental Medical Science
Lund University
Lund, Sweden
salma.kazemi_rashed@med.lu.se*

2nd Klara Esbo

*dept. Clinical Sciences
University of Gothenburg
Gothenburg, Sweden
klara.esbo@gu.se*

3rd Mariam Miari

*dept. Clinical Sciences
Lund University
Malmo, Sweden
mariam.miari@med.lu.se*

4th Malou Arvidsson

Lund University Lund, Sweden

5th Sonja Aits

*dept. Experimental Medical Science
Lund University
Lund, Sweden
sonja.aitis@med.lu.se*

Abstract—In numerous image-based studies, such as cell studies, a plethora of information is embedded in cell objects, which must first be accurately extracted before it can be utilized for further downstream analysis and linked to biological processes explored in the original experiment. One such technique for this purpose is instance segmentation. In our study, which cell death and associated pathways and regulatory networks are investigated through a genome-wide microscopic screen, the dataset is extensive and challenging to manually curate or annotate. Therefore, we opted to train distinct U-shape nuclei segmentation models using only a small but meticulously annotated subset of the high-content screen data to ensure high accuracy where the superior U-Net model achieved an F1-score of 89%. Subsequently, we have deployed the best U-Net model to segment whole objects in the nuclei channel of the original genome-wide dataset, comprising nearly 5 million image frames across 213 plates. The extracted features from the nuclei channel served as biological metrics to explore cell death. The results identified several well-known genes involved in established cell death pathways, such as apoptosis. Additionally, other genes were discovered that merit further investigation for their effects on altering the visual features of cell nuclei.

Index Terms—Cell death pathways, Image Segmentation, Genome-wide study, U-Net.

I. INTRODUCTION

Cell death is a biological process that varies in its mechanisms depending on the initiating factors. It is essential for life, as it allows multicellular organisms to create space for new cells and prevent the spread of potentially harmful mutated cells. Cell death is a highly regulated process. However, if the regulation genes become mutated this can result in either excessive or insufficient cell death both of which contributes to numerous diseases. Abnormalities in cell death regulation are thus a key cause of many health disorders. There are several types of cell death. It could be highly regulated process where DNA breaks down, the nucleus shrinks, and the cell divides

into small fragments or triggered by cell trauma or damage, which cause an uncontrolled process where the cell swells and bursts [1]. Other types of cell death can lead to distinct morphological changes in various parts of the cell, such as the nucleus, the cell body, and organelles like lysosomes. Lysosomes, which are responsible for breaking down cellular macromolecules, contain enzymes that can become harmful if released into the cytosol. When the lysosomal membrane becomes permeable (LMP), it may trigger either regulated or unregulated forms of cell death, depending on the degree of permeabilization. This often amplifies downstream cell death pathways and can also impact other organelles, such as mitochondria. In this study, our focus is primarily on the nuclei of cells. We aim to first extract morphological features, such as shape and number, and then link these features to the human genome to identify associated cell death pathways [2].

This goal is accomplished by analyzing our own dataset using deep learning-based U-shaped models to segment the objects. The data involved a substantial dataset of nearly 9 TB image data across more than 213 plates and initially lacked annotation. To tackle these issues, we developed a stepwise solution. First, we manually annotated a small subset of our own dataset as just 50 images [3]. Then, leveraging similar publicly available datasets, such as broad bioimage benchmark collection (BBBC), we trained basic U-Net models [4] from scratch and experimented with more advanced models featuring pretrained encoders, such as Hover-Net [5]. We evaluated the models on a small test set before applying them to the full dataset. The simplicity of our chosen model allowed us to achieve a balance between computational efficiency and performance, achieving a strong F1-score of 89% for our best model. We then applied our best-performing model across the entire dataset on the nuclei channel to segment each nucleus, extracting both size and count features. The analysis of these visual features in connection with the human genome revealed

numerous known genes involved in established cell death pathways, as well as previously unidentified genes that warrant further investigation. These novel findings could potentially uncover new pathways within our focus area of cell death.

II. DATA AND METHODOLOGY

A. Dataset

The full dataset consists of 4,276,224 image frames with the dimension of 1104x1104x3. The experiment was conducted by Dr. Sonja Aits as a gene knockdown, image-based study on bone cancer cells called U2OS osteosarcoma cells. The cells were transfected with a siRNA-library to silence 18,170 protein coding genes one by one in 384-well plate format as almost 213 plates. Where from each well field of view, 16 images were captured from the reactions of cells to the lack of that gene. The phenotypical changes that can arise in the process of knocking-down a gene can be visualized and measured from microscopy images. The images are taken in three different fluorescent channels by cells being labelled with three biological markers that characterize three different parts of cells. The nucleus was labelled by Hoechst 33342 staining where was the sole purpose of this manuscript. For each plate, a set of negative and positive cell death control columns were included. The negative controls included non-targeting siRNA (OTP), transfection agent only (Mock-transfected cells), and untransfected cells where the positive control columns were PLK1 and KIF11 siRNA and LIPA siRNA for lysosomal damage. A zoomed-in view of the plates and images, along with the three channels, is presented in Fig. 1.

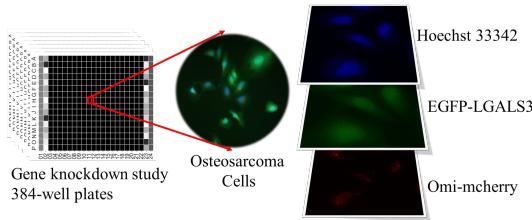


Fig. 1. This figure displays a sample of 384-well plate containing U2OS osteosarcoma cells, imaged across three distinct fluorescent channels.

B. Methodology

U-Net [4] is a U-shape convolutional neural network commonly used for image segmentation. It comprises a single down-sampling path or encoder, where spatial information is down-sampled using convolutional layers and pooling layers in order to reduce the resolution of feature map, retaining only high-level features (i.e., complex features of the input image, such as the objects to be segmented). In the second path of the model, known as the decoder, the layers increase the feature map's spatial resolution and up-sample it to the original image size to produce the segmentation map using skip connections from encoder. For extracting nuclei information from our own screen, We have trained the U-Net model from scratch with different set of images particularly a small subset of our own

images annotated meticulously by group members [3] (Fig. 2A).

After training U-Net models on different set-ups, we tested HoVer-Net [5], a network with an encoder and three decoder branches, used for histology image segmentation. HoVer-Net's decoder supports both instance segmentation and object classification, making it useful for segmenting heterogeneous environments with multiple cell types. It has three branches: the nuclear pixel (NP) branch for background identification, the HoVer branch for separating clustered cells, and the nuclear classification (NC) branch for classifying objects based on morphology and texture (which we did not use). We fine-tuned the HoVer-Net model with the same dataset used for U-Net models [3] (Fig. 2B).

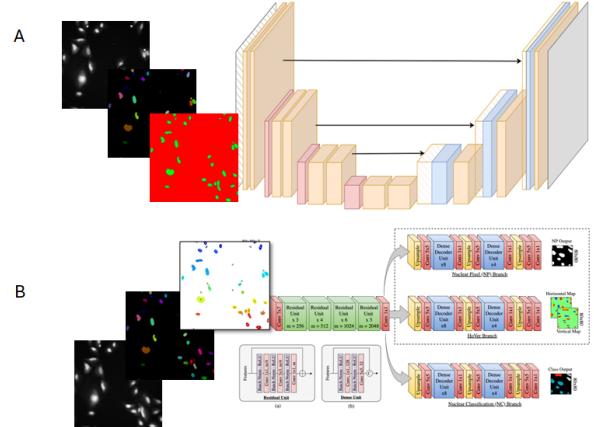


Fig. 2. A) U-Net: trained by original images alongside annotation maps that are reformatted into a three-class image label, classifying pixels as background, boundaries, or the interior class. B) HoVer-Net: trained with original images paired with annotation labels converted into ".mat" files, incorporating centric points added onto the segmentation map.

C. Metrics and evaluation

During training, models were evaluated on the validation set from [3], containing 10 annotated images from our dataset. Afterwards, for testing the generalize-ability of the best performing models, we took 3 models from each architecture and evaluated them on the test set from [3]. As main evaluation metrics we used F1-score at different Intersection-over-Union (IoU) thresholds and Jaccard Index. The IoU measures overlap of predicted and annotated objects and is calculated according to the equation 1.

$$IoU(A, B) = \frac{A \cap B}{A \cup B} \quad (1)$$

IoU determines Jaccard Index as a pixel-based index that can range from 0 to 1, from no overlap to identical objects. Since overlap of predicted and original objects is rarely 100% we selected a few different thresholds for comparison (50% to 90%). This is for calculating F1-score which is an object-based index. F1-score measures how many out of all predictions were true positives as the geometric average between the recall and precision. If a prediction is a true positive (sufficient overlap) is determined by the IoU threshold (90% in Fig. 3). Recall is

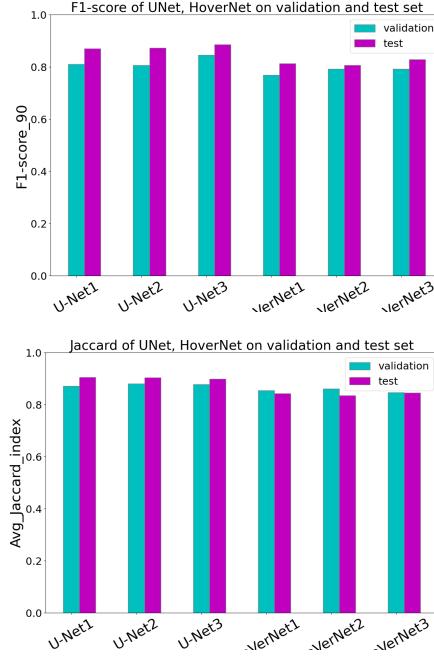


Fig. 3. This figure shows evaluation of U-Net and HoVer-Net models in terms of F1-score (threshold=90%) and average Jaccard Index on both validation and test set from [3].

a measurement of how many of the relevant objects that are retrieved, in this case nuclei, correctly found by the model.

After evaluation, U-Net3 was the best-performing model with the F1-score of 89% and Jaccard Index of 90% evaluated on test set. This was the model we applied over full dataset.

D. Results on full genome-wide data

After applying the best-performing model to the complete nuclei channel, we successfully segmented and quantified all nuclei across the entire dataset. The object counts and nuclei areas were normalized to the OTP control wells, and the resulting fold changes in nuclei counts and average area are visualized in Fig. 4. After analyzing the changes in nuclei numbers, we identified wells with a fold change of less than 0.5 and 0.1 and determined the corresponding knocked-down genes. This process resulted in a shortlist of candidate genes. We then examined which of these genes caused the greatest increase in average nuclei size. The overlapping genes between these two criteria are summarized in Table I and represent promising candidates for further investigation. This is because decreased nuclei count and swollen nuclei are both signs of cell death.

III. CONCLUSION

To analyze the genome-wide data, we began by annotating a small subset (approximately 0.009%) of the complete dataset. Using this annotated set, we trained models from scratch or leveraged transfer learning. After evaluating the models, we selected the one with the highest F1-score and applied it to the entire dataset. By identifying wells with the most significant

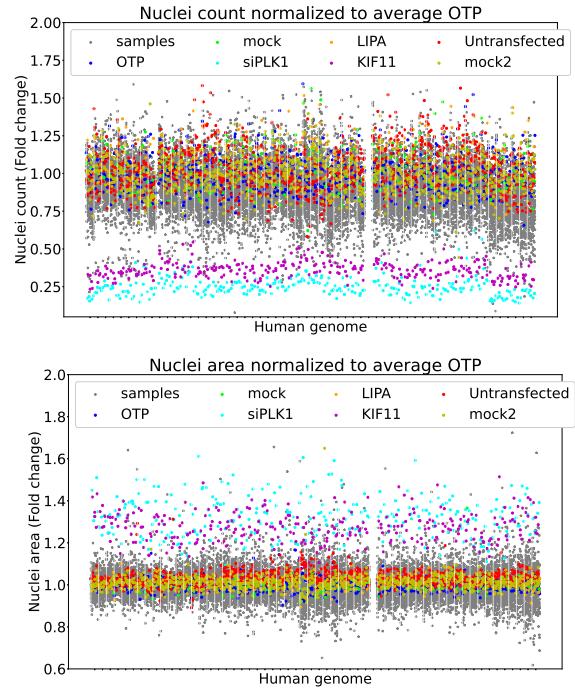


Fig. 4. The nuclei counts and nuclei areas are normalized to the OTP control wells, and the resulting fold changes in nuclei counts and average area are visualized.

TABLE I
GENES THAT SIGNIFICANTLY AFFECT COUNT AND AREA OF NUCLEI (FOLD CHANGE).

Entrez Gene Symbol	Nuclei Count	Nuclei Area
FBXO5	< 0.5	> 1.7
CCNA2	< 0.5	> 1.7
UBB	< 0.1	
POLR2A	< 0.1	

decreases in nuclei numbers and the largest increases in nuclei size, we linked these wells to the genes responsible for these changes. This process resulted in a list of candidate genes, whose roles in novel pathways of cell death will be further investigated.

REFERENCES

- [1] I. Bowen, *Cell death in biology and pathology*. Springer Science & Business Media, 2012.
- [2] M. Jäättelä and J. Nylandsted, “Methods for probing lysosomal membrane permeabilization,” *Cold Spring Harbor protocols*, vol. 2015, no. 11, pp. pdb-top070 367, 2015.
- [3] M. Arvidsson, S. K. Rashed, and S. Aits, “An annotated high-content fluorescence microscopy dataset with hoechst 33342-stained nuclei and manually labelled outlines,” *Data in Brief*, vol. 46, p. 108769, 2023.
- [4] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [5] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, “Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images,” *Medical image analysis*, vol. 58, p. 101563, 2019.