



HAL
open science

Deep learning for nuclear image analysis and application to 3D plant nuclei

Guillaume Mougeot, Frédéric Chausse, Sophie Desset, Katja Graumann

► To cite this version:

Guillaume Mougeot, Frédéric Chausse, Sophie Desset, Katja Graumann. Deep learning for nuclear image analysis and application to 3D plant nuclei. ORASIS 2021, Centre National de la Recherche Scientifique [CNRS], Sep 2021, Saint Ferréol, France. hal-03339633

HAL Id: hal-03339633

<https://hal.science/hal-03339633>

Submitted on 9 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deep learning for nuclear image analysis and applications to 3D plant nuclei

Guillaume Mougeot^{1,2}

Frederic Chausse¹

Sophie Desset³

Katja Graumann²

¹ Institut Pascal

² Oxford Brookes University

³ Génétique, Reproduction et Développement

guillaume.mougeot@uca.fr

Abstract

The continuous development of microscopy has led biologists to have access to large high-resolution 2D and 3D image datasets. Automatic analysis of cellular and nuclear images has become an important challenge in the bioimaging field. To help biologists extract information from these images, tools have been designed to count objects in the image, study object type, their localization or morphology. The current state-of-the-art is led by deep learning methods. Their development relies on the availability of large data sets, on enhancing graphical processing units (GPU) of computers and on developments of new methodologies such as convolutional neural networks (CNN). However, non-IT users may experience difficulties when trying to use these on their own images. This short paper contains our first results after reviewing the state-of-the-art methods in the domain. It first introduces the current difficulties when working with bioimages, then lists the existing datasets for nuclear images analysis. It then exposes some of the easy-to-use tools for bioimaging and points out the different problems related to their use. It finally presents a new dataset for 3D images of plant nuclei that is designed for benchmarking purposes. This results should shortly be published in a journal of biology.

Keywords

Deep learning, 3D segmentation, Confocal microscopy, 3D plant nucleus

1 Introduction

Studies the cell nucleus find many applications both in fundamental and applied research. One of the most noticeable by the number of publication on the topic is cancerous tissue analysis. As will be shown latter in this review, many of the research surrounding the nucleus is attached to classify and localise tumorous nuclei in histopathology images which are mostly 2D whole slide images stained with hematoxylin and eosin.[8]

To study role of chromatin and nuclear envelop proteins biologists have used mutants coupled with microscopy

imaging. Indeed, these proteins have a direct impact on the nucleus morphology which can be measured on microscopy images. This images are generally captured using bright-field microscopy, confocal microscopy or electron microscopy. To extract quantitative features from these images is an major step in any statistical analysis.

The current improvements in imaging techniques has drastically shorten the amount of time spent capturing high-resolution microscopy images. Consequentially, the amount of data generated by biologists has significantly increased and led them to seek for new methods leveraging the burden of handmade analysis. Automating the quantitative studies applied on their images would let them conducting important statistical analysis.

Concurrently, a variety of new techniques for image analysis has appeared in computer science, among which deep learning stands out. Deep learning is one of the most promising method, given its precision on many computer vision problems as classifying images, localising and extracting objects of interest or for image noise reduction.

Unfortunately, many problems arise when attempting to use this new methods for bioimaging. The first one is the amount of data required to make this type of method to work properly. Indeed, deep learning methods must be *trained* on large datasets manually annotated by human experts. Creating these sets of data might be very expensive and existing ones are often not publicly accessible. Another difficulty with deep learning is its configuration. In addition to a good computer graphic card, deep learning needs many software libraries to being setup beforehand. The difficulty increases steeply when it comes to reuse an implementation as deep learning libraries becomes obsolete very quickly and newer versions are often incompatible with older ones. Finally, deep learning codes are far from being boilerplate codes and almost always need fine configurations by an expert to work.

All deep learning methods are nowadays far from being accessible to biologists. However, some initiatives exist to help non-deep learning experts. The main goal of this review is thus to orient biologists to select the easiest meth-

ods to use and the most appropriate ones for their needs. A peculiar attention is attached to the methods vouching their accessibility both for an end user and a developer.

2 Easy-to-use deep learning tools for nuclear image analysis

To make a deep learning method accessible can be difficult due to the many components that need to be shared. We label a deep learning method as accessible if it provides the following components:

- A clear description of the model in the publication
- A code associated it
- A datasets for training and testing the model
- A documentation or some tutorials explaining the code
- One or several trained models on the datasets
- A complete description of all the configuration needed to reproduce the results in the publication. This must include all of the model hyperparameters configuration, as the number of training loop or the size of the batch of data input during training, as well as all the versions of the dependencies (the best is to provide a Dockerfile to quickly configure the code).
- an easy-to-use interface. The best when dealing with images would be to provide a Graphical User Interface, but due to the amount of work needed to develop such interface, a proper Command Line Interface might suffice.

Methods that provide all of the above criteria for nuclear image analysis are very rare. What is presented in a publication is often very different from reality. Unfortunately, biologists are often not aware of it, thinking that most of the tools are ready to use. In this section, we will try to focus on the few tools that are the easiest to use for a non-IT user looking for segmenting nuclei in a microscopy images. For a more exhaustive review we send the reader to our review article that will be published shortly or to this excellent work done by Caltech University[9].

The easiest deep learning methods are probably online tools which are ready to use without any setups. For nuclei segmentation task, one could use NucleAIzer [7]¹, CellPose [10]² or DeepCell [1]³. These three websites can be used by simply 'dragging and dropping' an image, pressing a 'start' button and downloading the resulting image. The NucleAIzer [7] website contains a model (an extension of the Mask R-CNN [6] model) trained on the 2018 Data Science Bowl dataset which contains segmented mammalian nuclei of several types: small or large nuclei, from colored or grayscale 2D images of different resolution. Cellpose

[10] has been designed for cell segmentation specifically but can be used for nuclei segmentation as well. The online tool only works in 2D but the offline version offers more feature and works on 3D images as well. DeepCell [1] is a similar solution to Cellpose. The online version let the user segment 2D images and the offline version also supports 3D segmentation as well as cell tracking in 2D videos. All three solutions are open-source. Their codes contain much more options than the ones presented in the online interface but are more complex to use for a non-IT user. Cellpose⁴ for instance provides an additional graphical user interface to see the segmentation results and to manually correct them to interactively improve the training accuracy of the model, a "human-in-the-loop process". DeepCell⁵ provides a tool for the cloud-based deployment of deep learning models: DeepCell Kiosk⁶. DeepCell and Cellpose finally provide a documentation to customize the input dataset, the deep learning models or the training process.

Online tools are easy-to-use as no setup is required and no complex configuration either. One of the main disadvantages is that a user has to upload the image on an unknown server which could be complicated with sensitive data. Another problem arises due to the size of the data needed to be uploaded because the online tools only accept single images with a limited size.

DeepImageJ [5] or U-Net [4] plugins are other example of available userfriendly tools. These plugins are integrated into ImageJ, a software frequently used by biologists.

DeepImageJ [5]⁷ is a host platform for deep learning models: several models are offered each working on a specific task on bioimages. A user can then download the most fitted model regarding its needs. Four types of tasks are listed on their website: segmentation, super-resolution, virtual staining and density map estimation. After installing DeepImageJ, a user can import the downloaded model into the plugin and apply it on one image. All of this procedure is transparent to the user and can be done directly via a Graphical User Interface (GUI). Note that most of the models need pre-processed images as input, a step usually done by normalizing them. This step can also be done directly from the GUI. Because each model has been trained on a specific type of images and object, good performance on a certain task will only be observed if the users images matches the images of the model training dataset. For example, the model called 'U-Net Pancreatic Segmentation' has been trained on segmenting pancreatic cells captured with phase contrast microscopy and thus might work well only on similar images. Additionally note that DeepImageJ can only be used on 2D images. Finally, for user who would like to incorporate their own trained model in DeepImageJ.

¹www.nucleaizer.org

²www.cellpose.org

³www.deepcell.org

⁴<https://github.com/MouseLand/cellpose>

⁵<https://github.com/vanvalenlab/deepcell-tf>

⁶<https://github.com/vanvalenlab/kiosk-console>

⁷<https://deepimagej.github.io/deepimagej/>

U-Net [4]⁸ authors also offer an ImageJ plugin. Contrary to DeepImageJ, U-Net is a specialized plugin for image segmentation only on one type of images. However, the plugin offers an easy way to fine-tune the model on more specific tasks with only few images. The default implementation provides a model for 2D and 3D image segmentation. The biggest disadvantage with this plugin is the computer setup. It requires a backend server to compute the segmentation, which can be difficult to install with the appropriate dependencies especially if the images are large and a setup for a Graphical Processing Unit (GPU) is needed. The setup can however be slightly facilitate by using Docker.

3 Datasets for nuclear image analysis

Data is a key component in deep learning method development. Major improvements in deep learning are based on the publication of new datasets. The nuclear image analysis field follows this rule. The main datasets that are worth mentioning are listed in Table 1. DeepCell, CellPose and NucleAIzer, previously mentioned were, for instance, all trained on the 2018 Data Science Bowl dataset[3]. As can be noted, most of them are focusing on 2D image segmentation from mammalian tissues. Very few datasets exists for 3D application and none for plant image analysis.

We thus decided to contribute to the field by creating a novel datasets for 3D images analysis of plant nuclei captured by a confocal microscope. This dataset is composed of 413 annotated 3D images. These annotations were produced using a semi-automated method thanks to the ilastik[2] software. Each image contains one single annotated plant nuclei and each pixels has an approximate xyz-resolution of (0.1,0.1,0.2) micrometers. One sample of this image dataset is displayed on Figure 1.

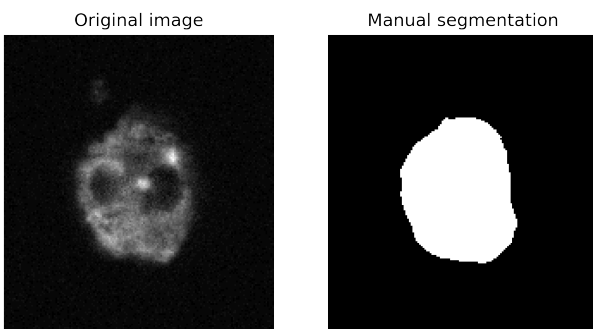


Figure 1: Sample of data from our dataset of 3D plant nuclei. **Left:** original raw image, **Right:** manually segmented nuclei

Images of plant nucleus are generally harder to segment than mammalian ones. The main reasons are the large variation in nuclei morphology (as opposed to mammalian nu-

⁸<https://lmb.informatik.uni-freiburg.de/resources/opensource/unet/>

clei which are often rounds), the noise in the image blurring the nuclei edges and the presence of a large nucleoli one of the cell constituents which are not stained and thus have the same color as the background.

Table 1: Available datasets for nuclear image classification and segmentation in 2D and 3D

Topic	Description
Nucleus classification 2D	Classification of nuclear atypia in breast cancer biopsy slides. Approximately 10400 frames. https://mitos-atypia-14.grand-challenge.org/Dataset/
Nucleus segmentation 2D	Multi-organ nuclei segmentation challenge. Challenge of MICCAI 2018. 30 images with approximately 22000 nuclear boundary annotations. https://monuseg.grand-challenge.org/Data/
Nucleus segmentation 2D	50 annotated histopathology images. https://zenodo.org/record/1175282
Nucleus segmentation 2D	Fluorescence microscopy images. 2661 cell nuclei of 37 fluorescence microscopy images http://www.cs.bilkent.edu.tr/~gunduz/downloads/NucleusSegData/
Nucleus segmentation 2D	A large variety nuclei images under a variety of conditions. Kaggle competition proposed by Booz Allen Hamilton https://www.kaggle.com/c/data-science-bowl-2018/data
Nucleus segmentation 2D	21000 nuclear boundaries in H&E-stained tissue. Used in many publications. https://nucleisegmentationbenchmark.weebly.com/dataset.html
Nucleus segmentation 2D	A comprehensive analysis of different studies on tissue images as nuclei segmentation, epithelium segmentation, tubule segmentation, lymphocyte detection, mitosis detection, invasive ductal carcinoma detection and lymphoma classification http://andrewjanowczyk.com/wp-static/
Nucleus segmentation 2D	52 images of clustered stained nuclei https://www.dropbox.com/s/9knzkp9g9xt6ipb
Nucleus segmentation 2D	100 images of clusted stained nuclei http://murphyliab.web.cmu.edu/data/2009_ISBI_Nuclei.html
Nucleus segmentation 2D	Grayscale images from multiple bright field microscopes. 4600 images and 26000 segmented cells https://edmond.mpg.de/imeji/collection/I45s16atmi6Aa4s1
Nucleus segmentation 2D	Annotated fluorescent nuclear images of different tissue origins. https://www.ebi.ac.uk/biostudies/files/S-B5ST265/dataset.zip
Nucleus segmentation and classification 2D	Biggest current dataset. From The Cancer Genome Atlas: 5060 whole slide tissue images from 10 cancer types (approximately 5 billion segmented nuclei) automatically segmented and quality controlled + 1356 manually segmented patches from the TCGA from 14 cancer types (10 same + 4 new) https://app.box.com/s/yd4pbndk2bxtnourzbpvopga8dczsnz/folder/99
Nucleus segmentation 2D and 3D	Database of various medical image analysis problems. Some of the image sets focus on nuclei segmentation in 2D and 3D. https://bbbc.broadinstitute.org/image-sets Some of the images set containing manually segmented nuclei: -Nuclei of U2OS cells in a chemical screen (2D): https://bbbc.broadinstitute.org/BBBC039 -Drosophila Kc167 cells (cells and nuclei outlined in 2D): https://bbbc.broadinstitute.org/BBBC007 -Human U2OS cells (out of focus) (2D): https://bbbc.broadinstitute.org/BBBC006 -Human HT29 colon-cancer cells (diverse phenotypes) (2D): https://bbbc.broadinstitute.org/BBBC018 -Murine bone-marrow derived macrophages (2D): https://bbbc.broadinstitute.org/BBBC020 -Nuclei of mouse embryonic cells (3D): https://bbbc.broadinstitute.org/BBBC050
Nucleus segmentation 3D	CA1 hippocampus brain region captured with electron microscopy https://www.epfl.ch/labs/cvlab/data/data-em/
Nucleus segmentation 3D	52 cells with stained nuclei captured with a LSMF http://opensegspim.weebly.com/download.html
Nucleus segmentation 3D	Confocal microscopy image of mouse embryo http://dx.doi.org/10.6084/m9.figshare.c.3447537.v1

4 Conclusion

We here presented some of the main tools developed to analysis images of cell nuclei. We pointed out the difficulty of using these tools for biologists as well as for computer scientist due to lacks in the method sharing process. This models often need to be re-trained, hence the importance of having access to a good dataset. We listed some of the main available datasets for nuclear image analysis and underlined the lack of data for 3D images as well as for plant nuclei. We finally presented our newly created datasets composed of 413 3D-plant nuclei. This dataset will be publicly available and will serve for benchmarking the existing deep learning tools. Due to its size, this dataset might also be used in the future to train novel deep learning

models onto the specific task of 3D nucleus segmentation. The ultimate goal is to make this tool accessible as much as possible to both biologists and deep learning community.

References

- [1] D. Bannon, E. Moen, M. Schwartz, et al. DeepCell Kiosk: scaling deep learning-enabled cellular image analysis with Kubernetes. *Nature Methods*, 18(1):43–45, 2021.
- [2] S. Berg, D. Kutra, T. Kroeger, et al. Ilastik: Interactive Machine Learning for (Bio)Image Analysis. *Nature Methods*, 16(12):1226–1232, 2019.
- [3] J. C. Caicedo, A. Goodman, K. W. Karhohs, et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nature Methods*, 16(12):1247–1253, 2019.
- [4] T. Falk, D. Mai, R. Bensch, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67–70, 2019.
- [5] E. Gómez-de Mariscal, C. García-López-de Haro, L. Donati, et al. Deepimagej: A user-friendly plugin to run deep learning models in imagej. *bioRxiv*, page 799270, 1 2019.
- [6] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):386–397, 3 2020.
- [7] R. Hollandi, A. Szkalicity, T. Toth, et al. nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. *Cell Systems*, 10(5):453–458, 2020.
- [8] D. Komura and S. Ishikawa. Machine Learning Methods for Histopathological Image Analysis. *Comput. Struct. Biotechnol. J.*, 16:34–42, 2018.
- [9] E. Moen, D. Bannon, T. Kudo, et al. Deep learning for cellular image analysis. *Nature Methods*, 16(12):1233–1246, 2019.
- [10] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021.