Using StarDist for CoNIC2022 Challenge

Marie-Claire Blache
PIC
UMR CNRS 7247 - INRAE 0085
Nouzilly, France
marie-claire.blache@inrae.fr

Benoît Piégu
ISLANDe
UMR CNRS 7247 - INRAE 0085
Nouzilly, France
benoit.piegu@inrae.fr

Bertrand Vernay
ICI
CNRS UMR 7104 - Inserm U 1258
Illkirch, France
vernayb@igbmc.fr

***Abstract*—Recents releases of pre-trained deep learning models such as StarDist and Cellpose have lowered the expertise threshold required to deploy deep learning tools in biomedical research. We describe in this manuscript our submissions to the CoNIC 2022 Grand Challenge for which we trained a StarDist model for both nuclei segmentation and cell type classification.**

***Index Terms*—StarDist, histology, colon tissue, nuclei segmentation, cell classification**

I. Introduction

A StarDist model was trained for both CoNIC 2022 Grand Challenge tasks by adapting the released code (Schmidt et al., 2018, Martin et al., 2020). Our purpose was to test the ease of deployment and performances of the available open-source tools without having to create a new model. We selected StarDist as it allows to train a network for both instance segmentation and object classification.

II. Data preprocessing and augmentation

A. Data preprocessing

The 4981 image patches 256x256 initially released by the coNIC Grand Challenge organisers where used for training the starDist model (Graham et al., 2021 2022). We applied the default image normalisation method embedded in the starDist code. See code at this url: https:/https://github.com/stardist/stardist/blob/master/examples/other2D/multiclass.ipynb.

B. Data augmentation

We applied the default data augmentation method embedded in the starDist code random performing rotations, flips, intensity changes and addition of random gaussian noise. See code at this url: https:/https://github.com/stardist/stardist/blob/master/examples/other2D/multiclass.ipynb.

III. Method description

The training set release for the CoNIC2022 Grand Challenge contained 4981 image patches (256x256 pixels) with the associated ground truth for both instance segmentation of all nuclei in each patch and the classification for each nucleus in one of 6 classes (Epithelial, Lymphocyte, Plasma, Eosinophil, Neutrophil and Connective tissue cells).

To establish a StarDist CoNIC baseline we first trained StarDist for 500 epochs with the default parameters on a 4781 image patches subset (85% Training set, 15% Validation), 200 images were kept for our own performance assessment with the default StarDist parameters (code available at this url: https://github.com/bvernay/CoNIC_2022).

For the Segmentation and Classification Task final submission, the model was trained on the full set of 4981 patches and We then trained StarDist for 500 epochs, on the full image patches dataset (85% Training set, 15% Validation)with a modification of the UNet network unet_n_filter_base: 256, default value for StarDist is 32. The unet_n_filter_base determines the number of convolution kernels (feature channels) for the first U-Net layer. The number of convolutions kernels is then doubled after each down-sampling layer.

IV. Post Processing

The probability and non-maximum suppression thresholds are optimised after the training using the original StarDist code(code available at this url: https://github.com/bvernay/CoNIC_2022)

V. Results

The models' performances were measured via the Grand Challenge evaluation protocol.All models'files: config.json, thresholds.json and weights_best.h5 are available at this url: https://github.com/bvernay/CoNIC_2022

A. Segmentation and Classification Task 1

CoNIC MAIIA Baseline Results (4781 patches)
Overall PQ = 0.5872, PQ+ = 0.5873, mPQ+ =0.3380
Components (con, eos, epi, lym, neu, pla) = 0.3698, 0.2971, 0.1120, 0.5603, 0.2402, 0.4491

CoNIC MAIIA Full Results (full dataset of 4981 patches and 256 entry convolution kernels)
Overall PQ = 0.6073, PQ+ = 0.6003, mPQ+ = 0.3397
Components (con, eos, epi, lym, neu, pla) = 0.3629, 0.3211, 0.0592, 0.5905, 0.2742, 0.4305

B. Cellular Composition Task 2

For task 2, our best performing model is a StarDist model trained with the default parameters on a 4781 image patches subset, 200 images were kept for our own performance assessment.

CoNIC MAIIA Baseline Results
Overall r2 = 0.4508, mae = 4.3601, maape = 0.3630

Cellular components r2 (con, eos, epi, lym, neu, pla) = 0.4915, 0.7576, 0.7299, 0.2006, 0.2070
Cellular components mae (con, eos, epi, lym, neu, pla) = 5.0, 0.3661, 9.6696, 4.1071, 2.9018, 4.1161
Cellular components maape (con, eos, epi, lym, neu, pla) = 0.5801, 0.1362, 0.2233, 0.4444, 0.3138, 0.4800

VI. Conclusions

We took advantages of the existing StarDist models extensively used in our respective light microscopy facilities. The model code is easily modifiable and allowed us "biologists" to evaluate the performance of a simple StarDist model CoNIC Grand Challenge without expertise in computer science.

VII. References

* Graham, Simon, et al. "Lizard: A Large-Scale Dataset for Colonic Nuclear Instance Segmentation and Classification." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

* Graham, Simon, et al. "CoNIC: Colon Nuclei Identification and Counting Challenge 2022." arXiv preprint arXiv:2111.14485 (2021).

* Schmidt, Uwe et al."Cell Detection with Star-Convex Polygons",Medical Image Computing and Computer Assisted Intervention - 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II,265-273 (2018),

* Martin Weigert, Martin et al. "Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy". The IEEE Winter Conference on Applications of Computer Vision (2020).