Self-Supervised Segmentation of 3D Fluorescence Microscopy Images Using CycleGAN

Alice Rosa[1], Hemaxi Narotamo[1] and Margarida Silveira[1]

Abstract— **In recent years, deep learning models have been extensively applied for the segmentation of microscopy images to efficiently and accurately quantify and characterize cells, nuclei, and other biological structures. However, typically these are supervised models that require large amounts of training data that are manually annotated to create the ground-truth. Since manual annotation of these segmentation masks is difficult and time-consuming, specially in 3D, we sought to develop a self-supervised segmentation method.**

Our method is based on an image-to-image translation model, the CycleGAN, which we use to learn the mapping from the fluorescence microscopy images domain to the segmentation domain. We exploit the fact that CycleGAN does not require paired data and train the model using synthetic masks, instead of manually labeled masks. These masks are created automatically based on the approximate shapes and sizes of the nuclei and Golgi, thus manual image segmentation is not needed in our proposed approach.

The experimental results obtained with the proposed CycleGAN model are compared with two well-known supervised segmentation models: 3D U-Net [1] and Vox2Vox [2]. The CycleGAN model led to the following results: Dice coefficient of 78.07% for the nuclei class and 67.73% for the Golgi class with a difference of only 1.4% and 0.61% compared to the best results obtained with the supervised models Vox2Vox and 3D U-Net, respectively. Moreover, training and testing the CycleGAN model is about 5.78 times faster in comparison with the 3D U-Net model. Our results show that without manual annotation effort we can train a model that performs similarly to supervised models for the segmentation of organelles in 3D microscopy images.

Clinical relevance— **Segmentation of cell organelles in microscopy images is an important step to extract several features, such as the morphology, density, size, shape and texture of these organelles. These quantitative analyses provide valuable information to classify and diagnose diseases, and to study biological processes.**

I. INTRODUCTION

A. Motivation

Cell detection and segmentation is critical for CAD (Computer-Aided Diagnosis) as it supports various quantitative analyses, including calculation of cell morphology, e.g., size, shape, and texture. This is essential, for instance, for the analysis, diagnosis, classification and grading of cancer [3].

However, segmentation of subcellular structures in microscopy images presents many challenges, such as the presence of digital noise and background clutter, blurring, variations in the size, shape, and intracellular intensity/heterogeneity of nuclei/cells, and these are often grouped together into clumps so that they sometimes touch and/or overlap [4]. In recent decades, several automated segmentation methods for microscopy images have been investigated that aim to overcome some or all of these challenges.

Deep learning-based models have been successfully applied to computer vision tasks, including microscopy image analysis, for instance, for nuclei detection and cell segmentation. One popular deep architecture is the Convolutional Neural Network (CNN). CNNs have been successfully applied for 2D nuclei segmentation [5]. For the specific task of 3D microscopy image segmentation the 3D U-Net model [1] has been widely used in various works.

However, these supervised deep neural network methods require large amounts of pixel-wise annotated data for training, which can be time-consuming and require expert annotation. To address this, in recent years several weakly supervised [6], [7] and self-supervised [8], [9] deep learning models that perform well on data with few or even no annotations have been proposed. Generative Adversarial Networks (GANs) have been used to improve the performance of these approaches, due to their ability to generate and translate data from one domain into another one. A popular GAN model is the CycleGAN [10], which learns to perform these translations using unpaired data. That is, it only requires examples of images from the source and target domains which do not need to be paired.

In [11] a two-stage pipeline, that does not require a manually annotated dataset, was proposed for nuclei segmentation in 3D microscopy images. In this approach, firstly synthetic masks were generated automatically. Thereafter, a modified CycleGAN model was trained and used to convert synthetic masks into synthetic microscopy images. Finally, the synthetic masks and synthetic images obtained with the CycleGAN model were used to train a modified 3D U-Net for segmentation. Although this approach does not require manual segmentation of 3D nuclei, it requires the training of two separate models to perform segmentation which increases the complexity of the system and adds time and computational costs. Furthermore, training of the 3D U-Net does not use the real microscopic images, only the synthetic ones.

A.R, H.N and M.S were supported by LARSyS - FCT Project UIDB/50009/2020. H.N was supported by the Fundação para a Ciência e a Tecnologia (FCT) Doctoral Grant 2020.04511.BD. We acknowledge Doctor Cláudio Franco for providing the dataset used in this work.

[1] Institute for Systems and Robotics (ISR/IST) LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal alicehrosa@tecnico.ulisboa.pt, hemaxi.narotamo@tecnico.ulisboa.pt, msilveira@isr.tecnico.ulisboa.pt

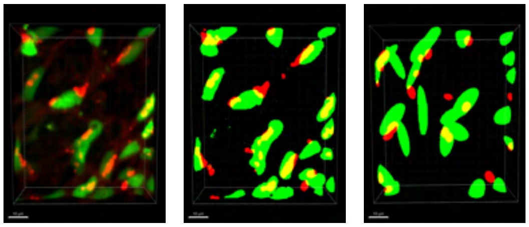
B. Objectives

In this work, we propose a one-step approach to segment two cell organelles in 3D microscopy images: nuclei and Golgi. Our approach is an extension to 3D of the 2D CycleGAN model [10]. We generate synthetic masks automatically based on the typical shapes and sizes of the cell organelles that we aim to segment. Thus, our approach also does not require the laborious manual annotation step. However, in [11] the authors create both the synthetic masks and synthetic microscopy images. In our proposed approach, we only generate the synthetic masks, and use these and the real images to train a CycleGAN model to perform segmentation directly. Thus, our proposed approach is much simpler compared to the segmentation method proposed in [11]. Additionally, we propose to use the original CycleGAN model for segmentation which is computationally less intensive compared to the modified CycleGAN used in [11].

II. METHODOLOGY

A. Dataset

The dataset used in this work consists of three 3D fluorescence microscopy images of mouse retinas (Fig. 1(a)). These images range in size from 1143x1010x55 to 2694x1981x61 [12]. In addition, manually labeled segmentation masks are used for training the supervised methods and evaluating the performance of all methods. In these masks the red and green channels contain the segmentation masks of the Golgi and nuclei, respectively (Fig. 1(b)). To use these images and masks to train the models, it is necessary to divide them into equal sized patches of size 64x64x64. To get the 64 slices in the z-axis for the images and ground-truth masks, reflection padding was applied.



(a) (b) (c)

Fig. 1. Examples of patches extracted from images and masks. (a) Microscopy Image (b) Corresponding Ground-truth Mask (c) Synthetically unpaired generated Mask.

Synthetic segmentation masks: In order to train the proposed self-supervised model, synthetic segmentation masks had to be created (Fig. 1(c)).

For this purpose, ellipsoids and spheres are created at random positions to represent the nuclei and Golgi, respectively. For each nucleus-Golgi pair created, the radius of the spheres (Golgi) and the size of the principal and secondary axis of the ellipsoids (nuclei), the distance between the nucleus and Golgi (relative to the center of the Golgi), and the rotation of each pair are selected randomly from the intervals [5,9] pixels, [[17,30],[8,13],[8,13]] pixels, [-6,6] pixels and [0,180] degrees, respectively. For each image created, a random number of nucleus-Golgi pairs were generated (between 45 and 70 pairs). To make the images more realistic, an elastic transformation was applied to each nucleus-Golgi pair. This transformation consists in deforming the image using displacement vectors and a spline interpolation. The value of this displacement was set to 0.75. Three synthetic masks, of the same size as the three microscopy images in our dataset, were created.

B. Proposed Approach

For this work, semantic segmentation of cell nuclei and Golgi will be performed using CycleGAN [10]. Two different domain images will be considered, the fluorescence microscopy images (domain I) and the synthetic segmentation mask (domain S). Like the original CycleGAN model, the architecture of this segmentation model will be composed of four interconnected networks, two generators (G_S and G_I) and two discriminators (D_S and D_I). A representation of this network can be found in Fig. 2. The generator (G_S) learns to map from domain I to domain S, and is used to obtain the segmentation ouput, after the network is trained.

The generator and discriminator architectures used in the original CycleGAN model are intended for a 2D image-to-image translation task. For our dataset, this approach needs to be adapted to perform a 3D image-to-image translation. This can be achieved by replacing the 2D convolutional layers of the original CycleGAN with 3D convolutional. The loss function used to train CycleGAN is the same as the one proposed in the original paper [10]. Our implementation is available at `https://github.com/alicerosa20/SS-Seg-CycleGAN`.

C. Comparison between different approaches

The performance of our proposed self-supervised CycleGAN model is compared with two well-known supervised models for segmenting nuclei and Golgi in fluorescence microscopy images, the 3D U-Net [1] and Vox2Vox [2], which is the 3D extension of the Pix2Pix GAN model [13].

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Segmentation Performance Metrics

Pixel-based metrics are used to evaluate the performance of the different models. Specifically, pixel wise precision, recall, and Dice Coefficient (DC) are used to evaluate these different approaches. These metrics are calculated separately for the objects to be segmented: nuclei and Golgi.

B. Data Pre-Processing

We applied contrast stretching to the microscopy images to improve the learning ability of the models by highlighting the contours of the objects and emphasizing the difference between the objects and the background. The values of the 95th and 98th percentiles are used as lower and upper bounds for this method. Additionally, we normalized the input images with values in the range [0,255] to values in the interval [0,1] by dividing each image value by 255.

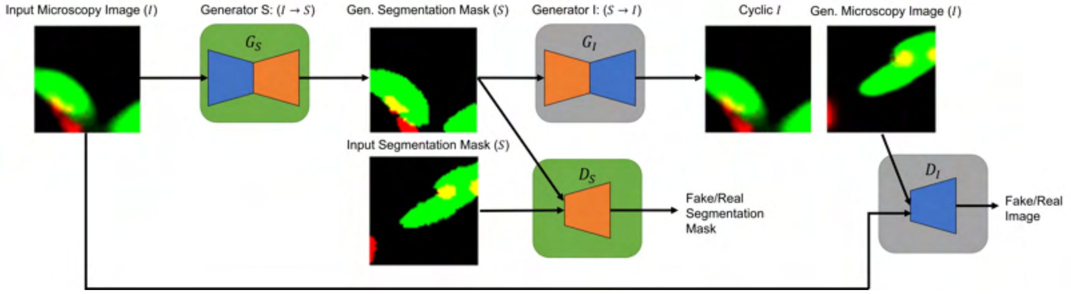

Input Microscopy Image (I)
Generator S: (I → S)
G_S
Gen. Segmentation Mask (S)
Generator I: (S → I)
G_I
Cyclic I
Gen. Microscopy Image (I)
Input Segmentation Mask (S)
D_S
Fake/Real Segmentation Mask
D_I
Fake/Real Image

Fig. 2. CycleGAN schematic for the proposed approach.

C. Implementation Details

All models were trained with a nested 3-fold cross-validation. Thus, for each approach, we trained 3 models with two microscopy images (one for training and one for validation) and evaluated the performance of each model on a held-out test microscopy image.

During training, the performance of the CycleGAN model is tracked by using the generator models to generate translated versions of a few randomly selected images at the end of each epoch. The model is stopped when it reaches collapse mode, i.e., when it produces exactly the same output image for different input images. After training, we visually analyse the results obtained by CycleGAN for different epochs and select the model from the epoch that gives better results for the validation set.

We changed the cycle consistency and identity loss to give more weight to the mean absolute error for Golgi (since it is the under-represented class). After some experiments, the weight given to this class was 3.5 and 1 to the nuclei class.

All experiments were performed in Python 3.9 on a computer with an NVIDIA GeForce GTX 1070 (8GB).

D. Execution Time

In this Section, we present the time needed to train the models. We estimated how long it takes to manually label nuclei and Golgi in a microscopic image. The estimate was 84 hours. For training time, the time needed to obtain these ground-truth masks was considered. In the case of the CycleGAN model, the time needed to create the synthetic masks (40 minutes) was added to the training time. The time needed to select the CycleGAN model that gives better results can be neglected. The final values are presented in Table I.

TABLE I
TIME NEEDED TO TRAIN EACH MODEL.

U-Net	Vox2Vox	CycleGAN
(84h × 2) + 66 min = 169.1 hours	(84h × 2) + 74 min = 169.2 hours	40 min + 28.6 h = 29.3 hours

E. Results and Discussion

Table II presents the results obtained for the segmentation of nuclei and Golgi with the 3 models implemented (average of cross-validation results). The best results for each class are highlighted in bold. Fig. 3 is a 3D visualization of the segmentation results obtained for two pre-processed microscopy image patches.

Segmentation of the nuclei class is challenging due to background clutter and low contrast in the images. The Vox2Vox model obtained the best Dice coefficient of 0.7947, followed by CycleGAN with 0.7807 and U-Net with 0.7756. The CycleGAN and Vox2Vox models had similar DC values, with CycleGAN being inferior by 1.4%. The CycleGAN model was sensitive to digital noise and had relatively low precision (0.7497), but was best at segmenting nuclei because it can segment nuclei even when they have low contrast (recall of 0.8433).

The challenges in classifying the Golgi class are mainly digital noise and small size of the Golgi, thus segmentation of these organelles is a difficult task. The best Dice coefficient was obtained for the U-Net model with 0.6834, followed by the CycleGAN model with 0.6773 and the Vox2Vox model with 0.5993. The U-Net model also obtained the best precision (0.7248) and CycleGAN the best recall (0.8001) for the Golgi class. Analysing the segmentation masks (Fig. 3) and quantitative results obtained for these models, we can conclude that, among the compared methods, U-Net provides the best segmentation masks of Golgi with less false positive pixels and the best Dice coefficient value. Moreover, CycleGAN is the model that segments the Golgi with less false negative pixels.

As mentioned earlier, supervised models are highly dependent on the data on which they are trained. The annotations are difficult to obtain and if they have errors or are imperfect, these imperfections are propagated to the models. We have found that some ground-truth masks have problems, such as not all objects being segmented and the existence of some discontinuities in the z-axis, which have a negative impact on the performance of the U-Net and Vox2Vox models. Moreover, these errors in the ground-truth masks also influence the quantitative results.

Finally, we concluded that with the CycleGAN model we were able to achieve comparable results to the supervised models, but with execution time about 5.78 times faster.

IV. CONCLUSIONS AND FUTURE WORK

In this work we proposed a self-supervised CycleGAN model for the segmentation of 3D microscopy images. From the experimental results, we concluded that with our CycleGAN model, we were able to obtain similar results to

TABLE II

AVERAGE PERFORMANCE METRICS VALUES OBTAINED FOR THE THREE DIFFERENT MODELS IMPLEMENTED FOR THE 3-FOLD CROSS-VALIDATION.

	DC Nuclei	DC Golgi	Precision Nuclei	Precision Golgi	Recall Nuclei	Recall Golgi
U-Net	0.7756	**0.6834**	0.8055	**0.7248**	0.7773	0.7353
Vox2Vox	**0.7947**	0.5993	**0.8079**	0.6886	0.8077	0.6829
CycleGAN	0.7807	0.6773	0.7497	0.6290	**0.8433**	**0.8001**

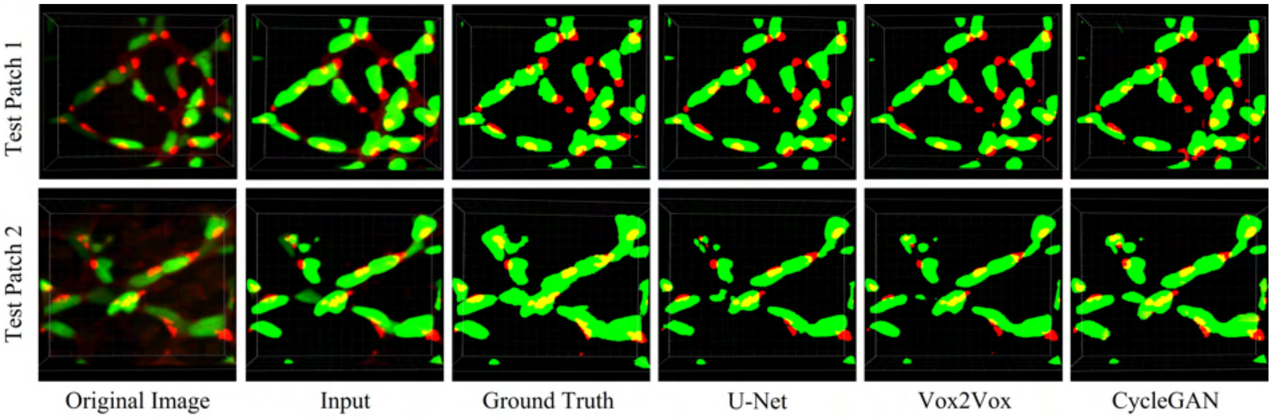


Fig. 3. Examples of test results for the segmentation of nuclei and Golgi.

the supervised models, with execution time approximately 5.78 times faster. It also has the advantage of being more transferable. However, it has the limitation that it has greater difficulty segmenting the smaller Golgi, and tends to oversegment because it has difficulty distinguishing background noise from actual nuclei and Golgi.

The proposed and implemented approaches have the limitation that they can only be used for semantic segmentation. Therefore, they are not able to distinguish between the different nuclei and Golgi and consequently, we are not able to, for example, count the different nucleus and Golgi in a microscopy image, which could then be applied in studies such as [12].

To segment individually each nucleus and Golgi in the image, future work could add a class in the nuclei and Golgi channels, to represent the borders of these organelles in the images, as done in [14] for 2D nuclei segmentation. Furthermore, this work can also be extended for instance segmentation, to distinguish the different nucleus and Golgi and in this way have more applications in the analysis of 3D microscopy images.

REFERENCES

[1] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *ArXiv*, vol. abs/1606.06650, 2016.

[2] M. D. Cirillo, D. Abramian, and A. Eklund, "Vox2Vox: 3D-GAN for Brain Tumour Segmentation," *CoRR*, vol. abs/2003.13653, 2020.

[3] T. Hayakawa, S. Prasath, H. Kawanaka, B. Aronow, and S. Tsuruoka, "Computational Nuclei Segmentation Methods in Digital Pathology: A Survey," *Archives of Computational Methods in Engineering*, vol. 28, pp. 1–13, 01 2021.

[4] F. Xing and L. Yang, "Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images: A Comprehensive Review," *IEEE reviews in biomedical engineering*, vol. 9, 01 2016.

[5] R. Hollandi, N. Moshkov, L. Paavolainen, E. Tasnadi, F. Piccinini, and P. Horvath, "Nucleus segmentation: towards automated solutions," *Trends in Cell Biology*, vol. 32, no. 4, pp. 295–310, 2022.

[6] H. Qu, P. Wu, Q. Huang, J. Yi, G. M. Riedlinger, S. De, and D. N. Metaxas, "Weakly Supervised Deep Nuclei Segmentation using Points Annotation in Histopathology Images," in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, vol. 102 of *Proceedings of Machine Learning Research*, pp. 390–400, PMLR, 08–10 Jul 2019.

[7] Z. Zhao, L. Yang, H. Zheng, I. H. Guldner, S. Zhang, and D. Z. Chen, "Deep learning based instance segmentation in 3D biomedical images using weak annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 352–360, Springer, 2018.

[8] D. J. Ho, C. Fu, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Segmentation of Fluorescence Microscopy Images Using Three Dimensional Convolutional Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 834–842, 2017.

[9] D. J. Ho, C. Fu, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei detection and segmentation of fluorescence microscopy images using three dimensional convolutional neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 418–422, 2018.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017.

[11] K. W. Dunn, C. Fu, D. J. Ho, S. Lee, S. Han, P. Salama, and E. J. Delp, "DeepSynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data," *Scientific reports*, vol. 9, no. 1, pp. 1–15, 2019.

[12] H. Narotamo, M. Ouarné, C. A. Franco, and M. Silveira, "Joint Segmentation and Pairing of Nuclei and Golgi in 3D Microscopy Images," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 3017–3020, 2021.

[13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, 2017.

[14] J. C. Caicedo, J. Roth, A. Goodman, T. Becker, K. W. Karhohs, M. Broisin, C. Molnar, C. McQuin, S. Singh, F. J. Theis, *et al.*, "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images," *Cytometry Part A*, vol. 95, no. 9, pp. 952–965, 2019.