Deep Learning-Based Instance Segmentation of Neural Progenitor Cell Nuclei in Fluorescence Microscopy Images

Gabriel Pérez[1,2(✉)], Claudia Cecilia Russo[2], Maria Laura Palumbo[1,3], and Alejandro David Moroni[1,3]

[1] Centro de Investigaciones y Transferencia del Noroeste de la Provincia de Buenos Aires (CIT NOBA)-UNNOBA-UNsADA-CONICET, Junín, Buenos Aires, Argentina
gabriel.perez@itt.unnoba.edu.ar, {mlpalumbo, admoroni}@comunidad.unnoba.edu.ar

[2] Instituto de Investigación y Transferencia en Tecnología (ITT), Universidad Nacional del Noroeste de la provincia de Buenos Aires (UNNOBA), Junín, Buenos Aires, Argentina
claudia.russo@itt.unnoba.edu.ar

[3] Centro de Investigaciones Básicas y Aplicadas (CIBA), Junín, Buenos Aires, Argentina

Abstract. In this work, a Deep Learning-based machine vision model was developed for the detection, segmentation and counting of Neural Progenitor Cell nuclei from fluorescence microscopy images. The cells were obtained from adult mice and cultivated in vitro, with cellular nuclei labeled using DAPI dye. Convolutional neural networks for instance segmentation, specifically the Mask R-CNN model with ResNet-50 and ResNet-101 backbones, were trained to recognize the nuclei, and their results were evaluated. Nuclei labeling was implemented semi-automatically, applying a Superpixel technique and then refining the segmentations from a manual process, also using a pre-trained model, which allowed to assemble a dataset of 66 images with 6392 labels in total. The results obtained with the Resnet-50 backbone show that there is an effectiveness of 98.6% for between the specialist count and model-predicted count, in addition to having an mAP50 of 98.0%. This approach has the potential to significantly reduce the time and effort required to analyze large image sets, which is especially useful in studies that require repetitive and detailed cellular analysis.

Keywords: Instance segmentation · Deep Learning · Fluorescence Microscopy · Cell Nuclei

1 Introduction

In the last few years, there has been significant progress regarding the acquisition of large sets of microscopy images in short periods of time, enabling computational analysis [1]. This paves the way for a series of tools capable of obtaining automatic information and supporting decision-making in the field of bioimaging [2, 3]. Some procedures carried out in laboratories are long, exhaustive, and repetitive, where the automation of different processes has become essential. In this context, digital image analysis plays an important role in studying patterns relevant to specialists [4].

M. Naiouf et al. (Eds.): JCC-BD&ET 2024, CCIS 2189, pp. 17–29, 2025.
https://doi.org/10.1007/978-3-031-70807-7_2

In the study of the hippocampus, a brain area related to learning and memory, there are Neural Progenitor Cells (NPC). These cells proliferate and then differentiate into three types of cells in the central nervous system: neurons, astrocytes, and oligodendrocytes. NPC from adult mice can be cultivated in vitro and studied through fluorescence microscopy images, which involves complex processes that could be automated using Deep Learning (DL) models.

Several approaches can be found in the literature about the automation through digital image analysis, for example, articles that provide a response to the characterization of cellular markers through algorithms based on color, texture, shape, and contrast [5, 6]. Likewise, there are works that do so through the use of Artificial Intelligence (AI) with supervised and semi-supervised learning. Various works use different Machine Learning (ML) techniques [7]. Particularly, Deep Learning (DL) models built with convolutional neural networks are used to obtain feature patterns of the elements to be recognized and then perform a classification process, involving general applications [8], cellular analysis [9, 10], and cellular tracking [11]. Among the techniques mentioned above, some focus on object segmentation [4, 11] and others focus on detection [11, 12]; both of which are an important first step for further analysis according to the specialist's needs.

ML methods are increasingly being implemented to characterize different markers in disease prognosis [13]. Through ML, disease biomarkers can be obtained without the need for a costly and invasive process, providing prognostic information similar to traditional techniques [2]. Specifically, there are pathology-dedicated softwares for the study of cell proliferation markers in renal cell carcinoma [3]. Image analysis alongside ML are tools of a great value for locating biomarkers; through a classifier, a heat map can be created to highlight tumor-infiltrating lymphocytes (TILs) and cancer cells data in an image, allowing for quantification controls in biomedical research and diagnosis [14].

In fluorescence microscopy imaging studies, fluorescent markers are commonly used to identify subcellular structures such as protein complexes, chromosomes, genes, and gene mutations [1]. To simplify the segmentation process, certain projects use fluorescent chromatin or DNA labels, where ML is widely used to classify cellular morphology obtained by fluorescent markers [15]. Particularly regarding the recognition and detection of cellular nuclei, there exist competitions like the Data Science Bowl 2018 [16], through which models like Mask-RCNN [17] could be trained to simultaneously detect and segment divergent nuclei from different sources to aid medical development and discovery.

An important step in the implementation of DL models is the assembly of the dataset and the corresponding labels. There are manual image labeling tools that allow to create bounding boxes, while others do so based on segments in the image, such as Superpixels [18]. It is indeed a laborious process that in certain areas of study can lead to errors if not done by the hand of a specialist in the field, potentially causing over-segmentation, underestimation, discard and add objects that are not of interest. These issues led to the research of new labeling methods based on models that automatically extract visual features, thus making the process easier. There exist articles where Superpixel models form an essential part of a classification process in medical images, such as in the

segmentation of brain MRI images [19], or in optic nerve images [20] where segments are used to extract relevant features.

The present research aims to develop a model based on vision and artificial intelligence to automate processes in genetic laboratories. This model, intended to obtain data and relevant information about cellular markers and behaviors, will be applied to a dataset provided by expert geneticists. Specifically, this article shows the development of a model with the ability to detect, segment, and count cell nuclei in an image using a convolutional neural network called Mask-RCNN [17]. To achieve this, fluorescence microscopy images of NPCs cultivated in vitro from adult mice were used, where cellular nuclei were labeled with DAPI dye [21]. Additionally, this work presents a semi-automatic approach for generating nucleus labels, through the use of a Superpixel and a proprietary pre-trained model, followed by a manual process.

2 Image Acquisition and Processing

2.1 About the Scene

Images were acquired by geneticists in the laboratory through a process that involved:

1. Extraction of Neural Progenitor Cells (NPC) from the hippocampus of adult BALB/c mice.
2. Cell cultivation in the presence of growth factors.
3. Cell proliferation analysis using immunofluorescence, which stained cellular nuclei with a DAPI marker in blue.
4. Acquisition of images using a Nikon Eclipse E800 fluorescence microscope with a 40× objective (Fig. 1).

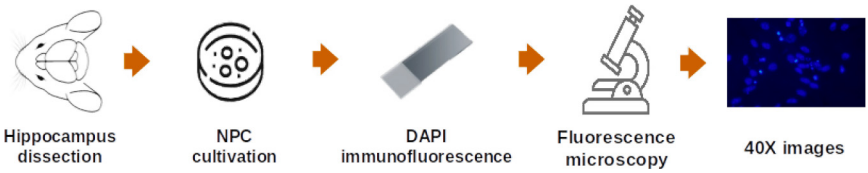

Fig. 1. DAPI image acquisition process. (Color figure online)

This process was performed by geneticists with the purpose of studying the hippocampus, an area of the brain related to learning and memory, where NPCs are found. These cells can proliferate and differentiate into three types of cells of the central nervous system: neurons, astrocytes, and oligodendrocytes. In order to study these behaviors, it was essential to analyze the presence of NPC nuclei through microscopic imaging.

In these images, we can see the presence of NPC nuclei, some of them separated and others overlapping. Additionally, there are small objects considered noise for the scene. The images have a resolution of 1280 × 960 pixels with the highest amount of data in the B (Blue) layer of the RGB color scheme.

2.2 Proposed Solution

The solution proposed in this work involves three stages, the first one consists in the preprocessing of DAPI cell nuclei images, where image enhancement techniques were used. The second stage consists of semi-automatic labeling of nuclei based on a Superpixel method and finally there is the training of an instance segmentation network of these nuclei.

Image Preprocessing and Enhancement. To begin with, the image was preprocessed in order to visually enhance it. The enhancement criteria were based on functions that can be applied to the image, such as contrast and brightness adjustment and noise attenuation. The purpose in mind was to highlight the objects of interest over the context in which they are located, thus creating contrast with other objects and the background.

In this step an enhancement was applied specifically in layer B (Blue) that composes the RGB of the color image, this was done because the image is purely blue and most of the data is in that layer. Contrast was achieved by applying a normalization technique only to this layer, considering its dynamic range, thus stretching the histogram (Fig. 2).

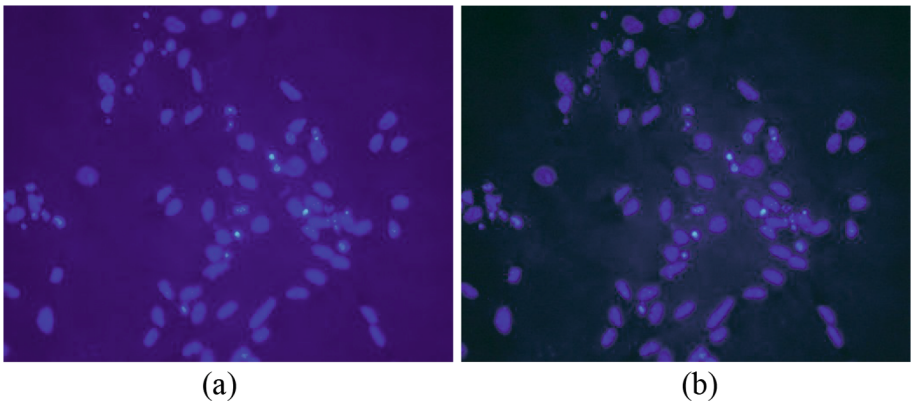

Fig. 2. (a) Original image. (b) Preprocessed image after normalization. The preprocessed image generates a contrast effect between the objects of interest and the background. (Color figure online)

Semi-automatic Labeling of Cellular Nuclei. On the preprocessed images, a set of labels was created to be used as input for the convolutional neural network. Given that the network to be used has the ability to detect objects in the image and also segment them, an approach to labeling that fits this capability was adopted. The inputs for this network are labels in segmentation or mask format, which is why this labeling methodology was approached.

Considering the characteristics of the image and the complexity of manually labeling cellular nuclei using polygons, a semi-automatic process based on a clustering algorithm was chosen. Specifically, the Felzenszwalb Superpixel algorithm [22] was used. This algorithm groups regions of the image that share similarities in terms of color and texture. As a result, several segments or superpixels were obtained that clearly separated the objects of interest from the surrounding background (see Fig. 3a).

To achieve a more effective isolation of the objects of interest from the background, it was considered that the background generally occupies the largest space in the image. Upon the set of segments or superpixels, the one with the largest area, corresponding to the background, was identified, and subsequently removed. This resulted in a clearer segmentation of the cell nuclei (see Fig. 3b).

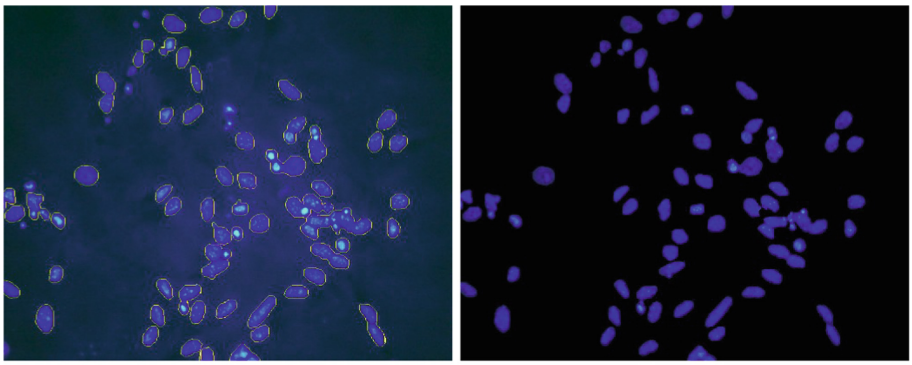

Fig. 3. (a) Preprocessed image of segments obtained from the Felzenszwalb Superpixel. (b) Image of segments obtained from Superpixel algorithm without the background. (Color figure online)

It is important to clarify that the algorithm relies on local contrast to identify the different elements in the image, that's why preprocessing was done before applying it. Figure 4 shows the result of applying it to the original image. It can be seen that there are areas where it couldn't detect some nuclei that are visible to the naked eye, this is due to the lack of contrast they present with the background.

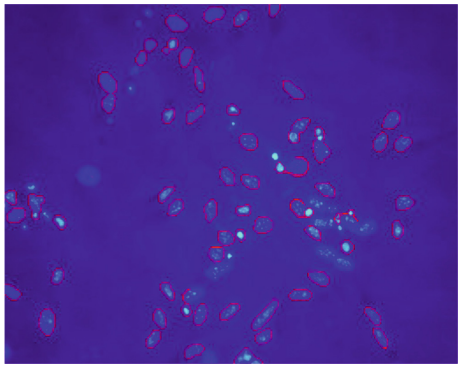

Fig. 4. Original image of segments obtained from the Felzenszwalb Superpixel.

Figure 3b shows that there are some superpixels that contain more than one cell nucleus. Object separation techniques were applied in these cases, but unfortunately,

the results were not satisfactory due to the morphological variations that some of these nuclei may present. In some cases, the different nuclei were separated appropriately, but the segmentation obtained was wrong (Fig. 5).

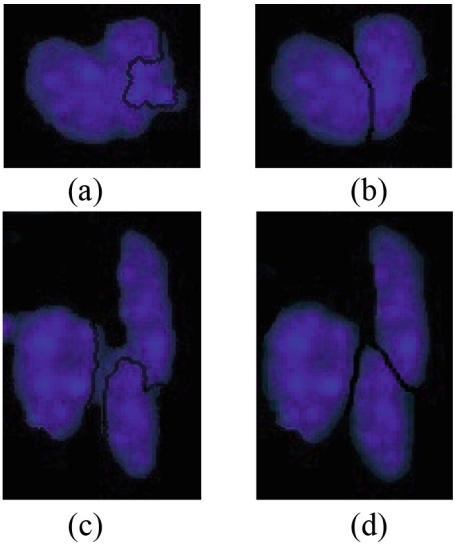

Fig. 5. Incorrect segmentation examples. (a) (c) Splitting with the Watershed algorithm. (b) (d) Desired splitting.

Considering that data is the raw material for deep learning models, incorrect segmentation does not faithfully represent the object. This can lead to errors in the training of the detection model and may distort the learning of a correct representation of the object. Based on this, it was chosen to manually separate connected superpixels of nuclei and remove segments that caused noise. This editing process was performed using a free-hand drawing tool on the segmentation images, eliminating incorrect parts and separating nuclei as necessary. Each labeled image was supervised by a geneticist and adjusted if necessary, ensuring proper segmentation.

Figure 6.a shows the final image with manually modified superpixels. Figure 6.b shows the input labels for the convolutional neural network, each cellular nucleus bounded by a contour representing the segmentation and additionally a bounding box locating it in the image.

Applied techniques were able to generate labels to train with diverse data under different illumination and contrast conditions. The final dataset from these techniques allowed to generate 22 original and 22 preprocessed images, both sets with 1773 labeled nuclei, resulting in a total of 3546 labels for the entire dataset.

Convolutional Neural Network Training. The network used for object recognition in this work is a two-stage convolutional network based on a Feature Pyramid Network (FPN) architecture. Specifically, the Mask R-CNN architecture [23] was employed, which is an extension of the Faster R-CNN model. In addition to object detection, Mask

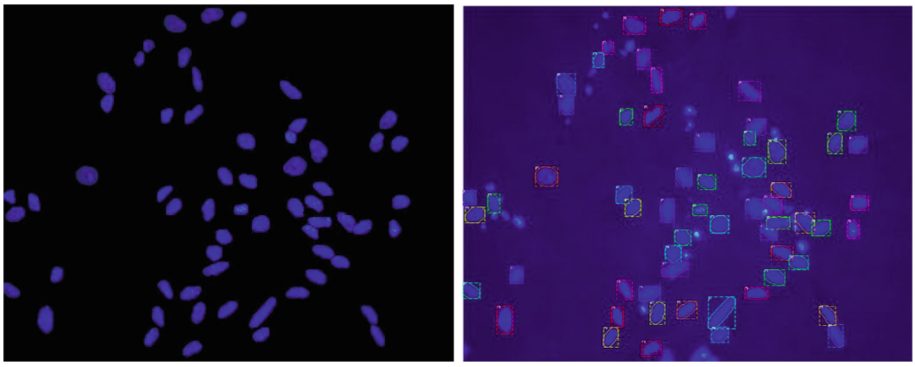

Fig. 6. (a) Image showing manually modified segments. (b) Image showing the input labels to the neural network.

R-CNN also provides segmentation of each detected object in the form of a binary mask. This technique, known as instance segmentation, allows to accurately distinguish the area of each object in the image, even in the presence of overlaps and occlusions.

Tests were performed on Mask R-CNN with ResNet-50 and ResNet-101 backbones [24]. ResNet-50 is a deep convolutional network with 50 layers that uses residual blocks to make training easier, while ResNet-101 is a deeper variant with 101 layers. The aim was to evaluate the impact of each of these backbones on this dataset, considering the difference in the network's ability to capture complex object features.

From preliminary tests using the semi-automatically obtained dataset and a Mask R-CNN neural network with ResNet-50, a model was obtained with the ability to recognize cell nuclei with an accuracy of 97.7%. Performance tests of the trained model were conducted with another dataset of 22 images, obtaining new labels to perform further training. In this case, the model was used as an automatic object labeller. Despite the high accuracy percentage, there were instances that went unrecognized or were misclassified. Therefore, small modifications were made to the model results, adding missing objects or removing incorrect ones, resulting in a new dataset.

The neural network training was then performed with the merging of the two datasets, resulting in a total of 66 images with 6392 labels. The images were split in groups of 75% for training and 25% for testing. All images used have both original and preprocessed versions as a data augmentation method to generate variability in the data environment. It is important to clarify that an original image and its preprocessed version are included in the same set, whether it's for training or testing, to minimize bias in both the training process and model testing, ensuring that there is no corruption between training and testing samples.

The original images have a resolution of 1280×980 pixels, which were resized to 1280×1280 using the square method before entering into the neural network. Training for both ResNet-50 and ResNet-101 was performed for a maximum of 50 epochs, with 100 iterations each, using pre-trained initial weights from the MS COCO dataset [25].

The best weights were selected based on the criterion of the lowest average percentage error between the actual count and the count predicted by the model on the test set.

The error was calculated as follows:

$$E = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{N^i_{detected} - N^i_{real}}{N^i_{real}}\right| \times 100\% \quad (1)$$

where:

n: the number of images in the dataset.

$N^i_{detected}$: the number of objects detected by the model in image i.

N^i_{real}: number of real objects in image i.

Additionally, metrics such as mAP IoU = 0.5 (mAP50), precision, recall, F1-score, and other errors between real and predicted quantities by the model were considered.

3 Results and Discussion

The results obtained in the evaluation of the models for recognizing cellular nuclei of NPCs can be observed in Fig. 7, where the results for both the ResNet-50 and ResNet-101 backbones are shown. The graphs display a scatter plot that relates the quantities obtained by the model and the visual count, along with the regression line showing the correspondence between both variables.

The exact results are shown in Table 1, where different model evaluation metrics are shown. The effectiveness of the model is calculated as 100% minus the average percentage error:

$$Ef = 100\% - \frac{1}{n}\sum_{i=1}^{n}\left|\frac{N^i_{detected} - N^i_{real}}{N^i_{real}}\right| \times 100\% \quad (2)$$

where:

n: the number of images in the dataset.

$N^i_{detected}$: the number of objects detected by the model in image i.

N^i_{real}: number of real objects in image i.

Other metrics related to the actual count of cell nuclei to those predicted by the model were also analyzed, as shown in Table 2. These metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R2), and also the maximum error between both counts.

The results of both architectures with the ResNet-50 and ResNet-101 backbones, show that in this context and for these images, a model with high depth of learning is not necessary. Cell nuclei exhibit similar characteristics in terms of texture, color, size, and shape, the challenge lies in the occlusion or overlap that they may present in different images. A complex model with a backbone like ResNet-101 may induce overfitting in this context.

It can be seen that both ResNet-101 and ResNet-50 have similar results in terms of mAP50 metrics and Effectiveness. This provides us with information about the results

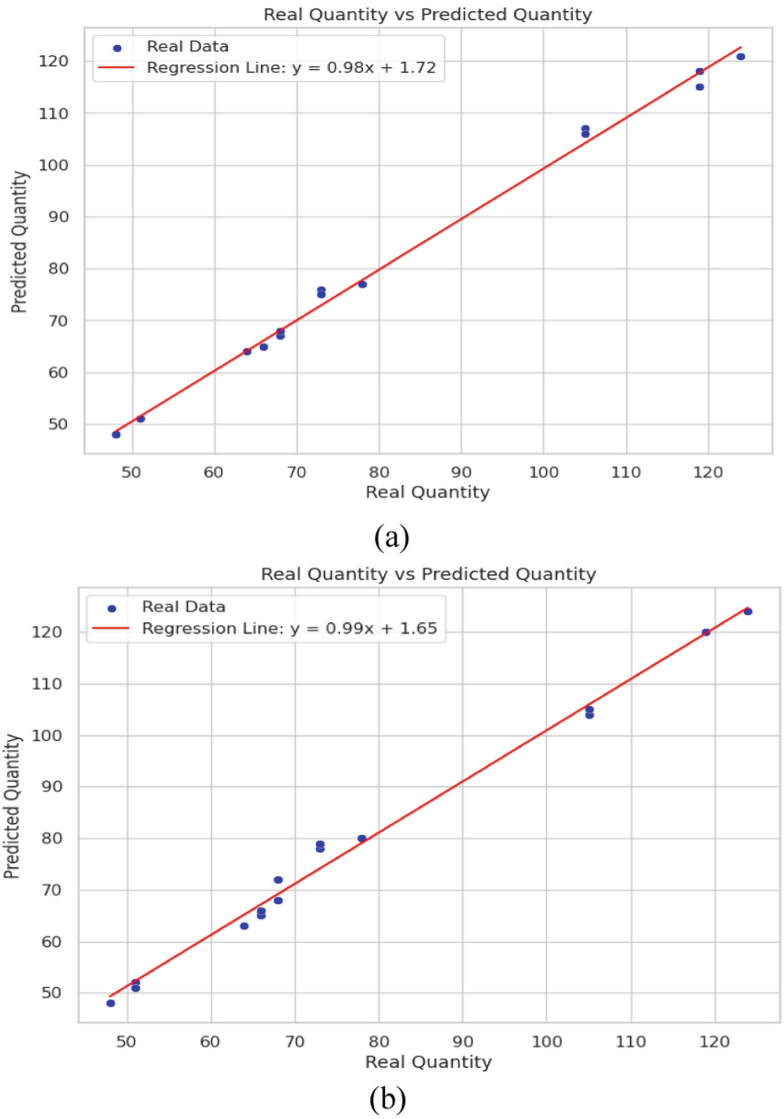

Fig. 7. Scatter plot relating quantities obtained by the algorithm and visual count, with the regression line showing the correspondence of both variables. (a) Resnet-50. (b) Resnet-101.

of the neural network itself as well as the counting performed by it. The major difference lies in the precision metrics and errors associated with the real count versus the count predicted by the model.

Despite the small differences in the mentioned metrics, it was found that ResNet-50 is the appropriate model for performing detections and segmentations of cell nuclei. This model not only yields the best results but is also lightweight and faster compared

Table 1. Model evaluation metrics.

Model	mAP50	Precision	Recall	F1-score	Effectiveness
Resnet-50	97.9	98.0	97.6	97.8	98.6
Resnet-101	97.7	95.9	97.4	96.7	98.0

Table 2. Model evaluation metrics based on quantities.

Model	MSE	RMSE	MAE	R^2	Max Error
Resnet-50	3.00	1.73	1.25	0.99	4
Resnet-101	5.44	2.33	1.44	0.99	4

to ResNet-101, which consumes more computational resources such as memory and processing time.

Regarding the results evaluated with ResNet-50 on the test set, we also analyzed the impact of including preprocessed images in the training and test sets.

Table 3. Resnet-50 evaluation metrics on original and preprocessed images.

Image type	Average Percentage Error	Effectiveness
Original	1.05	98.95
Preprocessed	1.64	98.36

Based on the results from Table 3, the model has higher effectiveness in recognizing original images, reaching nearly 99% (98.95%), whereas for the set of preprocessed images, the effectiveness is lower, at 98.36%. This provides us with the information that to achieve better results with our network, preprocessing before entering into the network is not necessary. It also indicates that the network has learned better from original images than from preprocessed ones.

In Fig. 8, the results of the model on an image of the same resolution but with a different number of nuclei are displayed. Figure 8.a has a real count of 51, where the model predicted 51, while Fig. 8.c has a real count of 105, and the model predicted 106.

Visually, it can be seen that the model developed in this work is considered important in aspects of detection, segmentation, and counting, which was already shown in the evaluation metrics where a mAP50 of 97.9% and an effectiveness of 98.6% between visual and predicted counts were obtained.

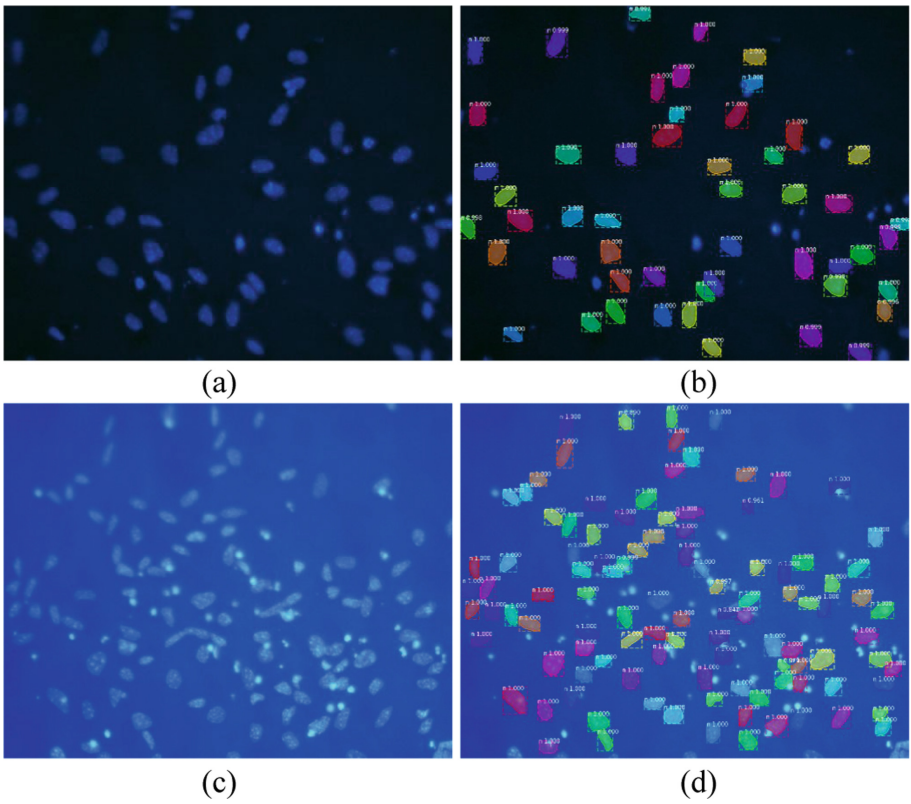

Fig. 8. Example of DAPI cell nuclei detections. (a) (c) Images. (b) (d) Model results.

4 Conclusion and Future Work

In this work, a Deep Learning-based artificial vision model was developed for the detection, segmentation, and counting of nuclei from Neural Progenitor Cells in fluorescence microscopy images. The proposed solution involved preprocessing and image enhancement stages, semi-automatic labeling of nuclei based on a Superpixel method, and training of an instance segmentation network for these nuclei.

The nuclei labeling was semi-automatic, images were preprocessed initially by applying filters and normalization to attenuate the noise and generate contrast between the objects of interest and the background. Then, Felzenszwalb Superpixel was applied to these images, whose segmentations required manual modification due to the presence of connected elements. This process allowed for the creation of a dataset of 44 images with a total of 3546 labels.

Preliminary tests using these labels and a Mask R-CNN neural network with Resnet-50 achieved a model capable of detecting cell nuclei with an accuracy rate of 97.7%. This model was used as an automatic object labeller in 22 new images, generating new labels. Despite the good performance of the model, certain false positives and false negatives had to be modified.

In the second stage, the neural network was trained by unifying the two datasets, resulting in a total of 66 images with 6392 labels. Tests were conducted on Mask R-CNN with Resnet-50 and Resnet-101 backbones to evaluate the impact of each backbone on this dataset. The model with the best results was obtained with Resnet-50, showing mAP50 metrics of 97.9% and an effectiveness of 98.6% between visual and predicted counts.

As future work, retraining of Mask R-CNN with new images verified by a geneticist specialist to improve its performance in result evaluation is planned, as well as adding different microscope sensors into the model. This work is part of a larger research project aimed at automatically analyzing the proliferation and differentiation behaviors of NPC cells. Future work will involve the analysis of cell proliferation, where the number of proliferating cell nuclei needs to be detected and counted against the total number of nuclei in the image. Additionally, cell differentiation analysis is planned, where the detection and distinction between each type of cell are required. In these cases, not only detections but also masks and counts of cell nuclei obtained by Mask R-CNN are crucial, as the presence of nuclei is essential to recognize different behaviors.

These results will contribute to the automation of laboratory processes, reduce the time spent on visual image analysis, generate decision support information, and provide tools for lab experts.

References

1. Allalou, A., Wählby, C.: BlobFinder, a tool for fluorescence microscopy image cytometry. Comput. Methods Progr. Biomed. **94**(1), 58–65 (2009). https://doi.org/10.1016/j.cmpb.2008.08.006
2. Harrer, S., Shah, P., Antony, B., Hu, J.: Artificial intelligence for clinical trial design. Trends Pharmacol. Sci. **40**(8), 577–591 (2019). https://doi.org/10.1016/j.tips.2019.05.005
3. Fuchs, T., Buhmann, J.: Computational pathology: challenges and promises for tissue analysis. Comput. Med. Imaging Graph. **35**(7–8), 515–530 (2011). https://doi.org/10.1016/j.compmedimag.2011.02.006
4. Xie, N., Li, X., Li, K., Yang, Y., Shen, H.: Statistical karyotype analysis using CNN and geometric optimization. IEEE Access **7**, 179445–179453 (2019). https://doi.org/10.1109/access.2019.2951723
5. Chadha, G., Srivastava, A., Singh, A., Gupta, R., Singla, D.: An automated method for counting red blood cells using image processing. Procedia Comput. Sci. **167**, 769–778 (2020). https://doi.org/10.1016/j.procs.2020.03.408
6. Haghofer, A., Dorl, S., Oszwald, A., Breuss, J., Jacak, J., Winkler, S.: Evolutionary optimization of image processing for cell detection in microscopy images. Soft Comput. (2020). https://doi.org/10.1007/s00500-020-05033-0
7. Kan, A.: Machine learning applications in cell image analysis. Immunol. Cell Biol. **95**(6), 525–530 (2017). https://doi.org/10.1038/icb.2017.16
8. Xing, F., Xie, Y., Su, H., Liu, F., Yang, L.: Deep learning in microscopy image analysis: a survey. IEEE Trans. Neural Netw. Learn. Syst. **29**(10), 4550–4568 (2018). https://doi.org/10.1109/tnnls.2017.2766168
9. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., Van Valen, D.: Deep learning for cellular image analysis. Nat. Methods **16**(12), 1233–1246 (2019). https://doi.org/10.1038/s41592-019-0403-1

10. Gupta, A., et al.: Deep learning in image cytometry: a review. Cytometry Part A **95**(4), 366–380 (2018). https://doi.org/10.1002/cyto.a.23701
11. Tsai, H., Gajda, J., Sloan, T., Rares, A., Shen, A.: Usiigaci: instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. SoftwareX **9**, 230–237 (2019). https://doi.org/10.1016/j.softx.2019.02.007
12. Mata, G., et al.: Automated neuron detection in high-content fluorescence microscopy images using machine learning. Neuroinformatics **17**(2), 253–269 (2018). https://doi.org/10.1007/s12021-018-9399-4
13. Carneiro, G., Zheng, Y., Xing, F., Yang, L.: Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis. In: Lu, L., Zheng, Y., Carneiro, G., Yang, L. (eds.) Deep Learning and Convolutional Neural Networks for Medical Image Computing. Advances in Computer Vision and Pattern Recognition, pp.11–32. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-42999-1_2
14. Klauschen, F., et al.: Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning. Semin. Cancer Biol. **52**, 151–157 (2018). https://doi.org/10.1016/j.semcancer.2018.07.001
15. Sommer, C., Gerlich, D.: Machine learning in cell biology – teaching computers to recognize phenotypes. J. Cell Sci. **126**(24), 5529–5539 (2013). https://doi.org/10.1242/jcs.123604
16. Kaggle.com. 2018 Data Science Bowl—Kaggle (2022). https://www.kaggle.com/c/data-science-bowl-2018. Accessed 11 Apr 2024
17. Johnson, J.W.: Adapting Mask-RCNN for Automatic Nucleus Segmentation. Advances in Intelligent Systems and Computing (2020). https://doi.org/10.1007/978-3-030-17798-0
18. Toro, C.O.: Algoritmos de segmentación semántica para anotación de imágenes. https://doi.org/10.20868/upm.thesis.55407
19. Li, H., Wei, D., Cao, S., Ma, K., Wang, L., Zheng, Y.: Superpixel-Guided Label Softening for Medical Image Segmentation. https://doi.org/10.48550/arXiv.2007.08897
20. Bechar, M., Settouti, N., Barra, V., Chikh, M.: Semi-supervised superpixel classification for medical images segmentation: application to detection of glaucoma disease. Multidimens. Syst. Signal Process. **29**(3), 979–998 (2017). https://doi.org/10.1007/s11045-017-0483-y
21. Kapuscinski, J.: Dapi: A DNA-specific fluorescent probe. Biotech. Histochem. **70**(5), 220–233 (1995). https://doi.org/10.3109/10520299509108199
22. Felzenszwalb, P.,Huttenlocher, D.: Efficient graph-based image segmentation. Int. J. Comput. Vis. **59**(2), 167–181 (2004). https://doi.org/10.1023/b:visi.0000022288.19776.77
23. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. https://doi.org/10.48550/arXiv.1703.06870
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. https://doi.org/10.48550/arXiv.1512.03385
25. Lin, T.-Y., et al.: Microsoft Coco: Common Objects in Context. https://doi.org/10.48550/arXiv.1405.0312