PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO

Título

SPERM CELL SEGMENTATION IN DIGITAL MICROGRAPHS BASED ON CONVOLUTIONAL NEURAL NETWORKS USING U-NET ARCHITECTURE

TRABAJO DE INVESTIGACIÓN PARA OPTAR EL GRADO ACADÉMICO DE MAGÍSTER EN INFORMÁTICA CON MENCIÓN EN CIENCIAS DE LA COMPUTACIÓN

AUTOR

Roy Kelvin Melendez Melendez

ASESOR

Dr. César Armando Beltrán Castañón

Junio, 2021

Sperm Cell Segmentation in Digital Micrographs based on Convolutional Neural Networks using U-Net Architecture

Roy Melendez*, César Beltrán Castañón†, Rosario Medina-Rodríguez‡
Pontificia Universidad Católica del Perú
Postgraduate School
Department of Engineering
Av. Universitaria 1801, San Miguel, Lima, Perú
*roy.melendez@pucp.edu.pe, †cbeltran@pucp.edu.pe, ‡r.medinar@pucp.pe

***Abstract*—Human infertility is considered a serious disease of the reproductive system that affects more than 10% of couples worldwide, and more than 30% of reported cases are related to men. The crucial step in evaluating male infertility is a semen analysis, highly dependent on sperm morphology. However, this analysis is done at the laboratory manually and depends mainly on the doctor's experience. Besides, it is laborious, and there is also a high degree of interlaboratory variability in the results. This article proposes applying a specialized convolutional neural network architecture (U-Net), which focuses on the segmentation of sperm cells in micrographs to overcome these problems. The results showed high scores for the model segmentation metrics such as precision (93%), IoU score (86%), and DICE score of 93%. Moreover, we can conclude that U-net architecture turned out to be a good option to carry out the segmentation of sperm cells.**

***Index Terms*—sperm cell micrographs, image segmentation, U-net architecture, deep learning**

I. Introduction

Infertility is a disease of the reproductive system; in men, this disease is diagnosed when sperm cells do not allow a woman to conceive a child or delay pregnancy after one or more years of regular unprotected sexual intercourse [1]. A crucial step in diagnosing male fertility is based on examining the morphology of the sperm cells, the main parts of which are the head, midpiece, and tail. In practice, the results derived from manual morphological analyses of sperm cells are highly dependent on laboratory technicians' experience. Furthermore, this manual evaluation is laborious, non-repeatable, time-consuming, and there is a high degree of interlaboratory variability.

Given this context, it is important to design accurate, automatic, and efficient artificial intelligence (AI) systems to improve the numerical analysis in sperm cell images. The morphology of sperm cells plays an essential role in the numerical analysis of these, which has aroused great interest in male infertility diagnosis. According to the World Health Organization (WHO) [2], there are abnormal categories in the morphology of sperm cells, which differ in shape, size, and texture in a very complicated way, making the task even more difficult for the expert [3].

This paper aims to implement and calibrate an advanced deep convolutional neural network architecture (U-Net) in the context of sperm cell segmentation. From the results, we can conclude that this architecture can segment microscopic images of sperm cells with good precision.

This article is organized as follows: In Section II, we review the research work in the area, focusing on scientific articles and commercial applications whose main objective is the segmentation of sperm cells. The materials and methods are described in detail in Section III, including the digital micrograph dataset, the U-Net architecture, our proposal, and the evaluation metrics. In Section IV, we present the experiments, results, and discussion. Finally, the conclusions can be found in Section V.

II. Related Works

For sperm cell segmentation tasks, conventional machine learning algorithms have been adopted to alleviate time-consuming work and improve segmentation performance. Although several approaches have been established to analyze sperm cells in animals, only a few approaches exist for the morphological segmentation of sperm cells in humans. We briefly review some approaches related to the morphological segmentation of sperm cells.

Yang [4] proposed a flagellum tracking algorithm to obtain flagellar movement patterns. Its framework aims to provide both head trajectories and flagellar movement patterns to assess sperm cell motility quantitatively. It distinguishes their work from other existing methods that analyze the motility of sperm cells based simply on the head's trajectories. Later, Van Raemdonck [5] implemented a computer-aided sperm cell analysis algorithm to accurately and efficiently analyze these cells' morphology. First, the authors remove the background, then cell segmentation is successfully used to sort cells into motile and immobile cells first, and motile cells sort into normal and abnormal cells second.

In other work, Medina et al. [6], presented an approach to sperm cell segmentation in micrographic images based on the combination of two well-known methods: background

subtraction based on the Lambertian reflectance model and morphological operations. The results showed a good performance in their set of images due to the presence of similar antecedents. However, the results also showed some issues in the segmentation due to the complexity of the images. Next, Mostajer [7] briefly reviewed the segmentation techniques of 2D grayscale images of sperm, and histogram-based thresholding algorithms are studied in detail. The authors propose that the combination of a non-linear transformation pre-processing with the histogram-based threshold for segmentation and detection of sperm cells.

In 2017, Chang [8] introduced a gold standard dataset, SCIAN-MorphoSpermGS, for the analysis and evaluation of the morphological classification of sperm cell heads. In particular, there was no open and free dataset available before this gold standard dataset was made public. It consists of 1854 images of sperm heads labeled by three Chilean experts in the reference domain specified by the WHO guidelines. Chang further proposed a two-phase analysis, CE-SVM, for the morphological classification of sperm cell heads in the SCIAN data set. In the first phase, a classifier is trained to distinguish the non-amorphous category from the remaining four categories. In the second phase, four classifiers are trained for the four amorphous categories, where each classifier aims to distinguish the specific non-amorphous category from the amorphous category.

From a different direction, Shaker [9] published the Sperm Cell Head Morphology (HuSHeM) data set and proposed an adaptive dictionary learning (APDL) based approach, which extracts certain square patches from the images of the head of these cells, to train dictionaries to recognize those categories of heads. In the evaluation stage, the square patches are recreated with the dictionary. The minimum general error among those of all the categories is calculated to identify the best head category. Recently, with the rapid development of deep learning techniques, Riordon [10] used a VGG16 (FT-VGG) architecture for the morphological classification of sperm cell heads.

Next, Nissen [11] used deep convolutional neural networks for the task of sperm cell segmentation and object detection. He was limited by computational time and precision demands in this task, which made training networks with many clustering layers difficult. In this work, we explore using full image training and upsampling of network outputs to increase performance to mitigate both issues. Afterward, Reza [12] presented a new framework for segmenting the head of sperm cells. The proposed method efficiently segments them into different shapes and sizes. Unlike classical learning methods that segment heads with residual parts of mid-pieces, their approach has precise detection and segmentation of sperm cell heads.

Finally, after reviewing the literature, we can realize that many of the investigations work on public data sets and with little noise, so their models operate without much difficulty and do not have an ideal architecture to perform the correct segmentation of sperm cells. For this reason, in this work, we used a single set of images where a considerable number of them have a strong presence of noise, different illumination conditions, and stains. Since calibrating an image processing algorithm is complicated, we propose a model based on convolutional neural networks using a U-Net architecture widely used in medical image segmentation.

III. Materials and Methods

Deep learning is a subtype of Machine Learning and belongs to the broader category of Artificial Intelligence, and such learning gets its name because it uses deeper networks compared to other Artificial Intelligence methods. One of the most common types is the Convolutional Neural Network, which is used for many computer vision tasks.

A. Dataset

The dataset contains 648 images in RGB format with size of 2452×2056 pixels (see Fig. 1). The images were acquired with an Axio Carl Zeiss optical microscope using $10x$ eyepieces and a $100x$ objective, in addition to an AxioCam ICc5 camera with $0.63x$ zoom. A considerable number of images in the dataset present a very noticeable noise, different degrees of illumination, and spots, making the segmentation process difficult with conventional machine learning methods; some examples of these images can be seen in Figure 2.

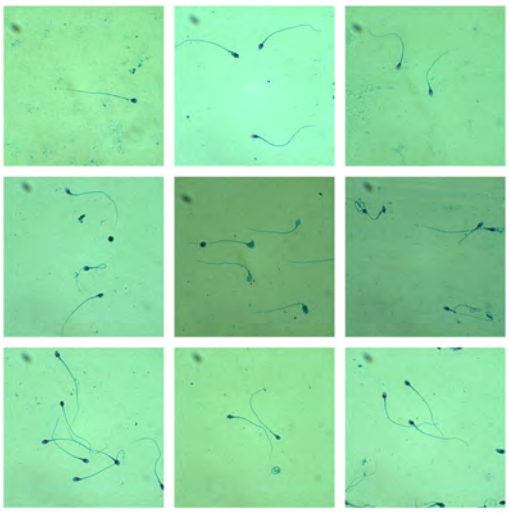

Fig. 1. Sample images of the dataset. As is visible, the images are high-resolution. It can also be seen that the shots were made with sharpness in a large number of these ones.

Moreover, as can be seen in Figure 3, we have binary masks that represent the ground-truth of the dataset, which were generated by first labeling each sperm cell in the image, next create a .json file, and finally, generate the binary masks for all the images in the dataset.

B. U-Net Architecture

This architecture makes its appearance in 2015, producing outstanding results in different segmentation problems with medical images. It is of the encoder-decoder type using a total

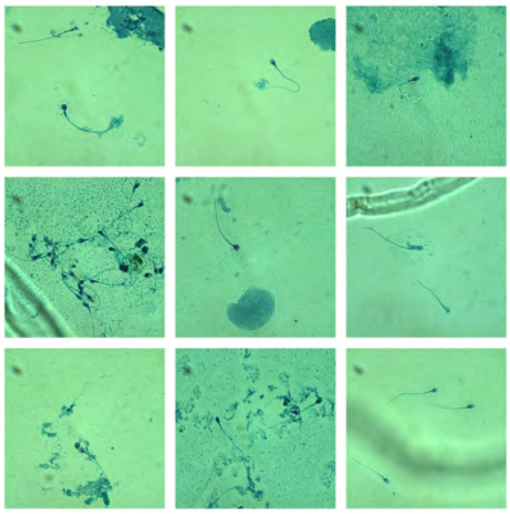

Fig. 2. Example of images from dataset with high presence of noise, different degrees of illumination and spots.

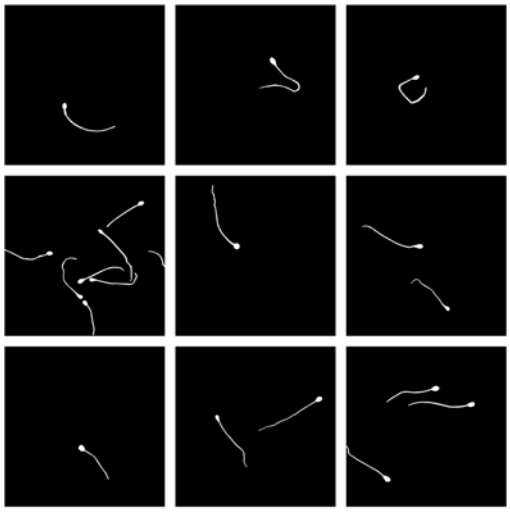

Fig. 3. Examples of Ground Truth Masks manually generated through a previous labeling in each sperm cell belonging from dataset.

of 23 convolutional layers, as can be seen in Figure 4. Its use is not restricted only to medicine because it can solve all kinds of segmentation problems [13].

The encoder stage is made up of blocks with max-pooling that reduces the resolution by half and double the channels. The decoder stage uses convolution transposed blocks that reverse the process. The more depth, the more significant the reduction, being $1/16$ of the proposed architecture entrance. The remarkable thing about this architecture is that it concatenates the input, before the reduction, with the output after the increase operation, managing to maintain the spatial information, which would otherwise have been lost.

As it can be seen in Figure 4, the symmetry of the architecture gives it a U-shape, which is why it is called U-Net. Its two parts are distinguished: the encoder on the left side and

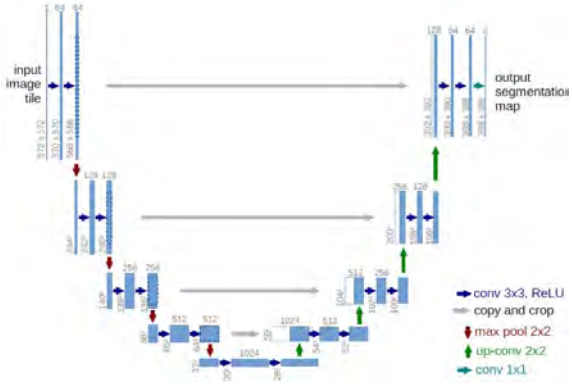

Fig. 4. U-Net Architecture. Figure extracted from [14].

the decoder on the right. They call the encoder the contraction path since the size is reduced due to the subsampling layers, and the expansion path decoder, in which the interpolations are carried out that progressively increase the size of the image. Likewise, five levels can be distinguished on each path that will be differentiated by their output size.

According to [15], among the advantages offered by the U-Net model for segmentation tasks, three important aspects stand out:

- The model allows the use of global location and context at the same time.
- U-Net works with very few training samples and provides better performance for segmentation tasks.
- The U-Net network architecture provides an end-to-end pipeline that processes the entire image in the forward pass and directly produces segmentation maps.

C. Sperm Cell Segmentation with U-Net

Given this context, we propose to work on a convolutional neural network architecture that has never been applied to sperm cell segmentation, which is the U-Net. As shown in Figure 5, convolution is the core of the U-Net architecture consisting of two 3×3 convolutions, each with a batch normalization and ReLU activation. All convolution layers in this function use the same number of filters.

Due to hardware limitations, the image input size is resized to 256 pixels, and the number of filters is $[16, 32, 48, 64]$ for each level. The decoder loops over the number of filters. Inside the loop, a 2×2 top sampling layer is applied. After the decoder part is completed, a 1×1 convolution is applied with the sigmoid activation function. Finally, it gives the final output in the form of a binary mask.

D. Evaluation Metrics

The Jaccard index [16] (also known as intersection over union, IoU) and the DICE similarity score [17], were used to assess the segmentation quality.

- Intersection over Union (IoU): represents the intersection of the image segmentation result and the ground-truth

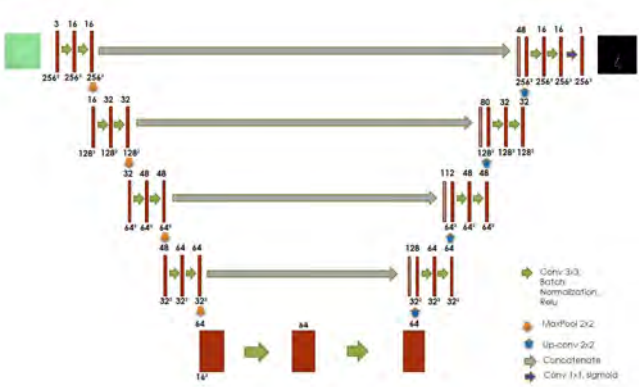

Fig. 5. U-Net architecture proposed in sperm cell segmentation. Where you can see the application of convolutions (3x3), followed by max-pooling and up-convolution operations in each filter (16, 32, 48, 64) corresponding to each level of the architecture.

divided by the union of the image segmentation result and the ground-truth, and is defined in Equation 1.

$$IoU = \frac{|SR \cap GT|}{(|SR| + |GT| - |SR \cap GT|)} \quad (1)$$

where SR is the result of the segmentation of the images and GT are the actual images of the original image tagging.

- The DICE similarity: is defined by Equation 2, this score compares the knowledge between the image segmentation result and the ground-truth.

$$DICE = \frac{2|SR \cap GT|}{(|SR| + |GT|)} \quad (2)$$

The ideal situation for the segmentation result is that SR and GT are completely coincident: IoU = 1, DICE = 1. The closer IoU and DICE are to 1, the better the image segmentation result.

IV. Experiments and Results

Different experiments have been carried out (see Table 1, Table 2, Table 3, Table 4), corresponding to hyperparameters settings established in the process of optimizing sperm cells segmentation. The dataset was divided as follows: 80% for training, 10% for validation, and 10% for testing. The best scores for IOU and DICE were achieved with a batch size of 4, a learning rate of 3e-4 and 30 epochs.

In these tables we can see how the results obtained in the segmentation of sperm cells have been progressively improved, in the first results we worked with a learning rate that did not help the network to achieve good segmentations in our architecture, for the stages following and after many tests, it was possible to find an adequate and optimal learning radius for our architecture, showing good segmentations as a result, already in the last stages of the experimentations, the data augmentation technique is applied to our dataset, from this way obtaining substantial improvements in our segmentations.

Data augmentation is the artificial generation of data employing disturbances in the original data, which allows us to increase both in size and diversity of our training data set. In computer vision, this technique became a standard for regularization, combat overfitting, and improve the performance of CNNs.

In our experiments, we obtained five more images for each original image in the dataset (see Fig. 6) by applying the following augmentation techniques: (i) CenterCrop crops an image from the center, which gives equal padding on both sides vertically and horizontally; (ii) RandomRotate90, lets the image rotate 90 degrees randomly; (iii) GridDistortion, generates a degree of distortion in the image with a given probability; (iv, v) HorizontalFlip and VerticalFlip, flips the given image horizontally and vertically, respectively; randomly with a given probability.

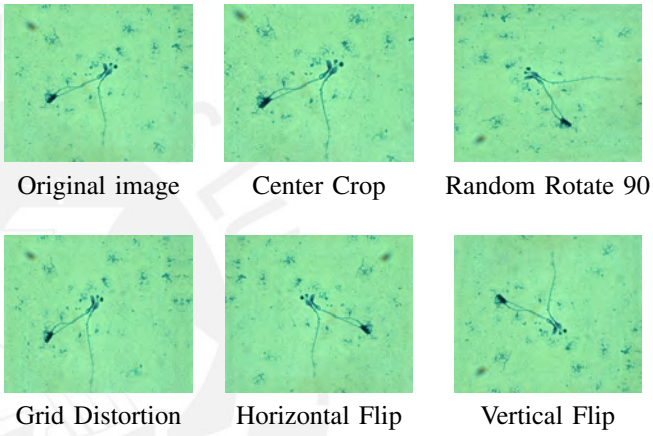

Fig. 6. Example of the data augmentation technique applied to our sperm cells dataset.

Thanks to these techniques, the dataset could be increased from 648 to 3888 images. The process also allowed to generate the ground truth corresponding to each image obtained through the same technique. Additionally, it allowed to significantly improve the segmentation coefficients IoU and DICE, which achieved better scores, thus allowing us to obtain better results in the segmentation of our sperm cells.

The experimental results were then compared with the annotated images (ground-truth), as shown in Figure 7. It can be seen that the segmentation results to our model after applying data augmentation results in better scores in both the IOU and the DICE, unlike a segmentation without applying this technique.

The evaluation indicators, IoU and DICE, were used as results of the quantitative analysis. Figure 9 shows the IoU and DICE curves of the network model training process; it can be seen that IoU and DICE are stable after approximately 30 epochs.

In Table I, we can also observe that in the test set, IoU reached a score of 66% and the DICE a score of 80%. On the other hand, in Table II we can see that in the test set, IoU reached a score of 88% and the DICE a score of 93%, which are a good scores for our problem. Both scores were achieved

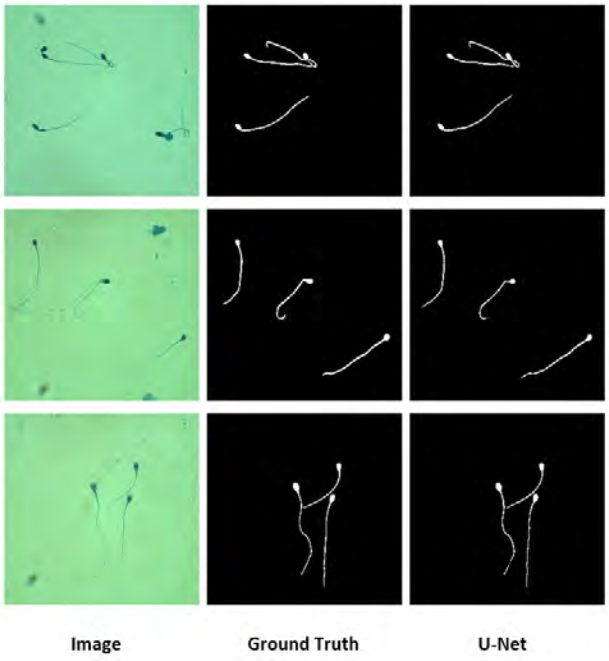

Fig. 7. Comparison with the predictions obtained through the U-net architecture in the dataset without data augmentation

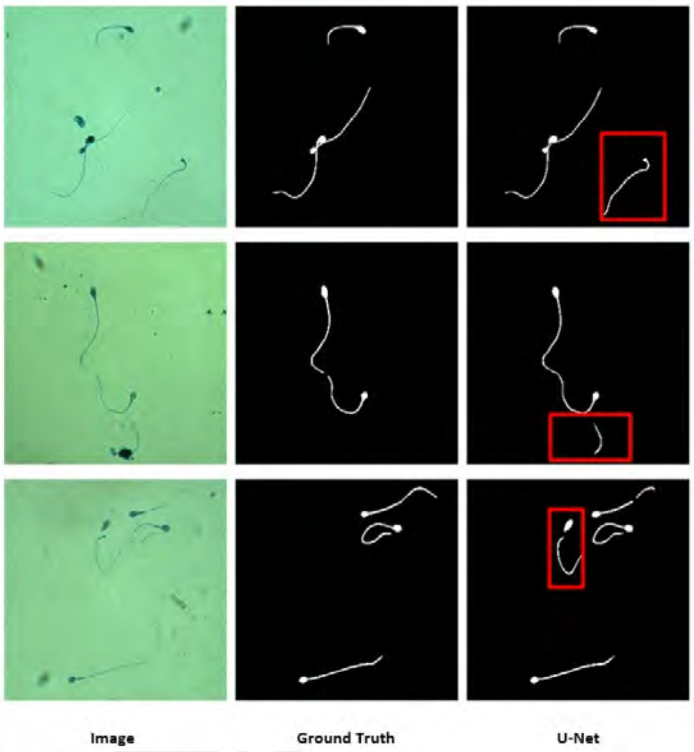

Fig. 8. Comparison with the predictions obtained through the U-net architecture in the dataset with data augmentation

due to a gradual adjustment of the hyper-parameters and data augmentation techniques. In the test set we can notice that for our model it is already being easy to learn, since we can see that the model is recognizing in the images those shapes that resemble a spermatic cell and we can see this in Figure 8. Therefore, it can be concluded that the results improved when using the data augmentation technique.

Finally, the loss function used was binary cross entropy, observing both the training data and the test data tend to converge as shown in Figure 10 . Therefore, we note that it is not necessary to overfit the data. Thus achieving a precision of 93% in the training set and 86% in the test set, considered to be good scores for our work, these can be seen in Table III and Table IV.

TABLE I
SEGMENTATION RESULTS OF DIFFERENT TRAINING EXPERIMENTS FOR THE DATASET WITHOUT DATA AUGMENTATION

		IoU			DICE		
Hyper-parameters		***Train***	***Val***	***Test***	***Train***	***Val***	***Test***
Epochs	20						
Batch	4	0.19	0.18	0.18	0.31	0.30	0.30
Lr	3.00E-4						
Epochs	30						
Batch	4	0.45	0.31	0.33	0.62	0.48	0.50
Lr	3.00E-4						
Epochs	40						
Batch	4	0.59	0.46	0.47	0.75	0.63	0.64
Lr	3.00E-4						
Epochs	50						
Batch	4	0.66	0.62	0.60	0.80	0.77	0.75
Lr	3.00E-4						

It worth mentioning that in a small number of images, the

TABLE II
SEGMENTATION RESULTS OF DIFFERENT TRAINING EXPERIMENTS FOR THE DATASET WITH DATA AUGMENTATION

		IoU			DICE		
Hyper-parameters		***Train***	***Val***	***Test***	***Train***	***Val***	***Test***
Epochs	10						
Batch	4	0.58	0.56	0.54	0.73	0.71	0.70
Lr	3.00E-4						
Epochs	20						
Batch	4	0.69	0.63	0.61	0.82	0.77	0.75
Lr	3.00E-4						
Epochs	30						
Batch	4	0.86	0.82	0.80	0.93	0.90	0.89
Lr	3.00E-4						

TABLE III
PERFORMANCE METRICS WITHOUT DATA AUGMENTATION

		Precision			Recall			F1-Measure		
Hyper-parameters		***Train***	***Val***	***Test***	***Train***	***Val***	***Test***	***Train***	***Val***	***Test***
Epochs	20									
Batch	4	0.82	0.68	0.69	0.63	0.67	0.65	0.71	0.67	0.68
Lr	3.00E-4									
Epochs	30									
Batch	4	0.87	0.81	0.88	0.79	0.51	0.52	0.83	0.64	0.65
Lr	3.00E-4									
Epochs	40									
Batch	4	0.92	0.79	0.80	0.85	0.69	0.68	0.88	0.74	0.74
Lr	3.00E-4									
Epochs	50									
Batch	4	0.92	0.81	0.81	0.86	0.71	0.69	0.89	0.76	0.75
Lr	3.00E-4									

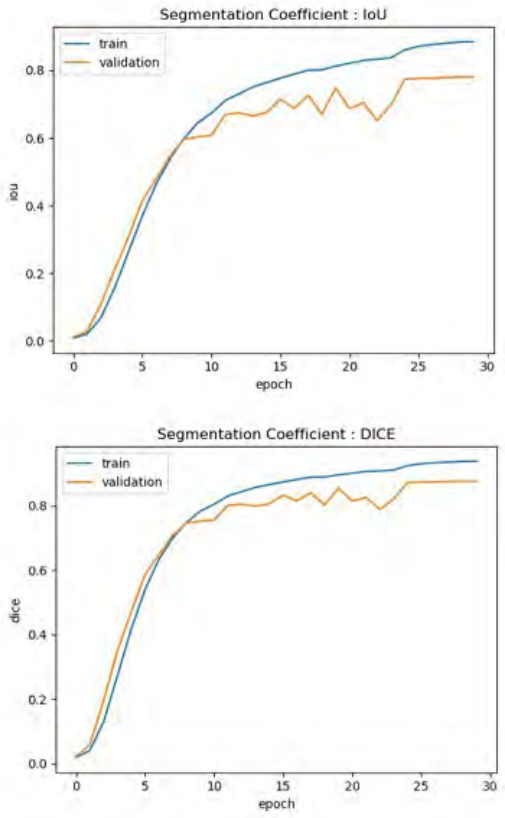

Fig. 9. Comparison between the IoU and DICE curves of the network model in the data set in a total of 30 epochs.

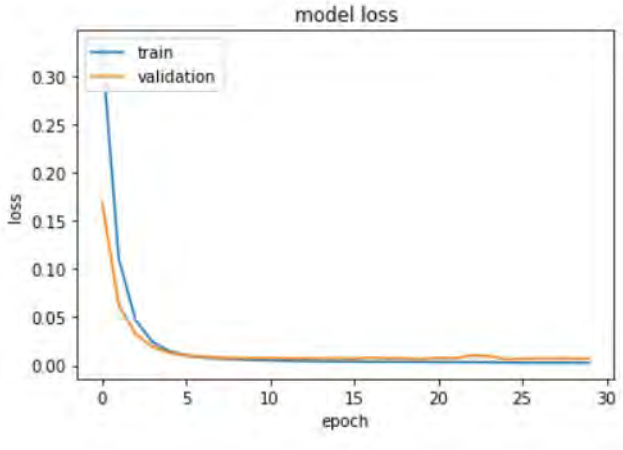

Fig. 10. Graph of Binary Cross Entropy Loss Function

segmentation did not occur so effectively due to the rolled shapes of the sperm cells, low illumination, and presence of spots (see Fig. 11) around 11% of the images obtained as a result of segmentation using the U-Net architecture did not obtain the results as expected, the remaining percentage (89%) achieved excellent segmentation results.

TABLE IV
PERFORMANCE METRICS WITH DATA AUGMENTATION

		Precision			Recall			F1-Measure		
Hyper-parameters		***Train***	***Val***	***Test***	***Train***	***Val***	***Test***	***Train***	***Val***	***Test***
Epochs	10									
Batch	4	0.85	0.83	0.82	0.78	0.72	0.72	0.81	0.77	0.77
Lr	3.00E-4									
Epochs	20									
Batch	4	0.91	0.86	0.85	0.83	0.76	0.75	0.87	0.81	0.80
Lr	3.00E-4									
Epochs	30									
Batch	4	0.93	0.87	0.86	0.86	0.77	0.76	0.89	0.81	0.81
Lr	3.00E-4									

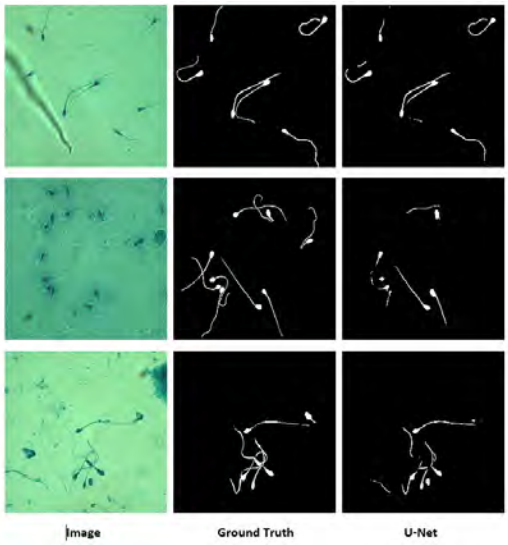

Fig. 11. Set of examples of the proposed segmentation, which results have some deficiencies.

V. CONCLUSIONS

The results obtained suggest that the U-Net architecture presents a good option for sperm cell segmentation tasks. As shown on the values obtained in both IoU and DICE, these are very acceptable for segmentation in sperm cells. Additionally, with this architecture, it has also been possible to demonstrate that it is much easier to segment images captured under different light conditions with noise and spots. It is worth noticing that a deep learning technique such as a convolutional neural network was applied to a dataset that no one has manipulated.

This model could also be improved by appropriately modifying the hyper-parameters used during the training process. On the other hand, it would be advisable to consider other techniques for increasing data, which help to improve the segmentation of sperm cells. As future works, we consider working with more filters in the U-Net architecture to see how it could improve segmentation in sperm cells.

REFERENCES

[1] V. Blasco, F. M. Pinto, C. González-Ravina, E. Santamaría-López, L. Candenas, and M. Fernández-Sánchez, "Tachykinins and kisspeptins in the regulation of human male fertility," *Journal of Clinical Medicine*, vol. 9, no. 1, p. 113, 2020.

[2] J.-C. Lu, Y.-F. Huang, and N.-Q. Lü, "Who laboratory manual for the examination and processing of human semen: its applicability to andrology laboratories in china," *Zhonghua nan ke xue= National journal of andrology*, vol. 16, no. 10, pp. 867–871, 2010.

[3] F. Zegers-Hochschild, G. D. Adamson, J. de Mouzon, O. Ishihara, R. Mansour, K. Nygren, E. Sullivan, and S. Van der Poel, "The international committee for monitoring assisted reproductive technology (icmart) and the world health organization (who) revised glossary on art terminology, 2009," *Human reproduction*, vol. 24, no. 11, pp. 2683–2687, 2009.

[4] H.-F. Yang, X. Descombes, S. Prigent, G. Malandain, X. Druart, and F. Plouraboué, "Head tracking and flagellum tracing for sperm motility analysis," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 310–313.

[5] L. E. Van Raemdonck, M. L. Davila-garcia, L. Mihaylova, R. F. Harrison, A. Pacey *et al.*, "An algorithm for morphological classification of motile human sperm," in *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2015, pp. 1–6.

[6] R. Medina-Rodríguez, L. Guzmán-Masías, H. Alatrista-Salas, and C. Beltrán-Castañón, "Sperm cells segmentation in micrographic images through lambertian reflectance model," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2015, pp. 664–674.

[7] F. M. Kheirkhah, H. S. Mohammadi, and A. Shahverdi, "Histogram non-linear transform for sperm cells image detection enhancement," in *2016 Eighth International Conference on Information and Knowledge Technology (IKT)*. IEEE, 2016, pp. 25–30.

[8] V. Chang, A. Garcia, N. Hitschfeld, and S. Härtel, "Gold-standard for computer-assisted morphological sperm analysis," *Computers in biology and medicine*, vol. 83, pp. 143–150, 2017.

[9] F. Shaker, S. A. Monadjemi, J. Alirezaie, and A. R. Naghsh-Nilchi, "A dictionary learning approach for human sperm heads classification," *Computers in biology and medicine*, vol. 91, pp. 181–190, 2017.

[10] J. Riordon, C. McCallum, and D. Sinton, "Deep learning for the classification of human sperm," *Computers in biology and medicine*, vol. 111, p. 103342, 2019.

[11] M. S. Nissen, O. Krause, K. Almstrup, S. Kjærulff, T. T. Nielsen, and M. Nielsen, "Convolutional neural networks for segmentation and object detection of human semen," in *Scandinavian Conference on Image Analysis*. Springer, 2017, pp. 397–406.

[12] R. A. Movahed and M. Orooji, "A learning-based framework for the automatic segmentation of human sperm head, acrosome and nucleus," in *2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2018, pp. 1–6.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[14] O. Ronneberger, "U-Net: Convolutional Networks for Biomedical Image Segmentation," https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/, 2015, [Online; accessed 08-January-2021].

[15] J. Tang, Y. Lan, S. Chen, Y. Zhong, C. Huang, Y. Peng, Q. Liu, Y. Cheng, F. Chen, and W. Che, "Lumen contour segmentation in ivoct based on n-type cnn," *IEEE Access*, vol. 7, pp. 135 573–135 581, 2019.

[16] A. Rosebrock, "Intersection over Union (IoU) for object detection," https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/, 2016, [Online; accessed 08-January-2021].

[17] J. Jordan, "An overview of semantic image segmentation." https://www.jeremyjordan.me/semantic-segmentation/, 2018, [Online; accessed 08-January-2021].