Deep Learning-Based Classification and Segmentation of Sperm Head and Flagellum for Image-Based Flow Cytometry

Paúl Hernández-Herrera[1,3], Victor Abonza[1],
Jair Sánchez-Contreras[1], Alberto Darszon[2],
Adán Guerrero[1]

[1] Universidad Nacional Autónoma de México,
Laboratorio Nacional de Microscopía Avanzada,
Mexico

[2] Universidad Nacional Autónoma de México,
Departamento de Genética del Desarrollo y Fisiología Molecular,
Mexico

[3] Universidad Autónoma de San Luis Potosí,
Facultad de Ciencias,
Mexico

{victor.abonza, darszon, adan.guerrero}@ibt.unam.mx,
paul.hernandez@uaslp.mx, jair.sanchez@im.unam.mx

Abstract. Image-Based Flow Cytometry (IBFC) is a potent tool for the detailed analysis and quantification of cells in intricate samples, facilitating a comprehensive understanding of biological processes. This study leverages the ResNet50 model to address IBFC's object-defocusing issue, an inherent challenge when imaging a 3D object with stationary optics. A dataset of 604 mouse sperm IBFC images (both bright field and fluorescence) underpins the exceptional capability of the ResNet50 model to reliably identify optimally focused images of the sperm head and flagella (F1-Score of 0.99). A U-Net model was subsequently employed to accurately segment the sperm head and flagellum in images selected by ResNet50. Notably, the flagellum presents a significant challenge due to its sub-diffraction transversal dimensions of 0.4 to 1 micrometers, resulting in minimal light intensity gradients. The U-Net model, however, demonstrates exceptional efficacy in precisely segmenting the flagellum and head (dice = 0.81). The combined ResNet50/U-Net approach offers significant promise for enhancing the efficiency and reliability of sperm analysis via IBFC, and could potentially drive advancements in reproductive research and clinical applications. Additionally, these innovative strategies may be adaptable to the analysis of other cell types.

Keywords. Deep learning, sperm, segmentation, classification, image-based flow cytometry.

1 Introduction

1.1 Image-Based Flow Cytometry in Sperm Analysis

Image-Based Flow Cytometry (IBFC) allows for quantitative exploration of cellular and subcellular characteristics such as morphology, protein distribution, and organelle localization [5].

These capabilities facilitate investigations into diverse processes like cell differentiation, cell-cell interactions, and disease-related alterations [2]. IBFC amalgamates flow cytometry, optical microscopy, and computational analysis [7].

Its capacity to rapidly capture thousands of images, positions it as an instrumental tool for uncovering and comprehending complex biological phenomena.

In the analysis of spermatozoa, there is a growing demand for IBFC due to its versatility in conducting multiparametric studies on various conditions and capturing the inherent population heterogeneity of these gametes.

This has significant implications in reproductive biology, medicine, as well as in agriculture and fisheries [6]. However, a limitation inherent to IBFC is its use of stationary optics to image tridimensional pseudo-stationary objects, resulting in a collection of images ranging from sharply focused to blurred (out-of-focus) [7].

A sharply focused sperm image, which we will refer to as the focused image for the rest of the paper, corresponds to an image where the sperm head and flagellum are clearly visualized. This focused image was obtained when the sperm is exactly in the focal plane of the objective lens [22].

On the other hand, if the sperm strays from the focal plane, which we will refer to as the out-of-focus image for the rest of the paper, a defocused or blurred sperm image forms. Analyzing an out-of-focus image may result in misinterpretations of the actual size, morphology, or intensity properties of the observed specimens [22].

1.2 Deep Learning Network

Deep learning is a subfield of machine learning that utilizes artificial neural networks [23] as a foundation to extract information and make predictions from data.

It has emerged as the leading approach for tasks like image classification and segmentation [11] and has demonstrated impressive outcomes across various domains, including sperm images.

In our study, we specifically investigated two networks: the ResNet50 for classification and the U-Net for segmentation of multi-channel images.

ResNet50, introduced by He et al. [9] in 2015, derives its name from "Residual Network" due to its innovative residual block, which sets it apart from conventional neural networks.

While traditional networks stack layers sequentially, ResNet50 incorporates residual connections that enable information to bypass specific layers, resulting in a more direct flow of information. This innovative structure tackles the problem of vanishing gradients, allowing for training with increased depth without suffering from performance degradation.

The ResNet50 architecture consists of 50 layers, including convolutional layers, pooling layers, and fully connected layers. As input data progresses through these layers, convolution and pooling operations are applied to extract significant features from the image. These features are then combined and utilized for classification.

ResNet50 has demonstrated outstanding performance in numerous classification tasks (Hossain et al. [10], Alnuain et al. [3], Anand et al. [4]), establishing new benchmarks in accuracy and robustness.

The U-Net2D architecture, introduced by Ronneberger et al. [19] in 2015, has a U-shaped architecture, consisting of an encoding stage and a decoding stage. In the encoding stage, convolutional and pooling layers are employed to extract meaningful features from the input image.

As the layers go deeper, the receptive field expands due to the pooling layers, enabling the capture of details at different scales. The decoding stage is responsible for reconstructing the segmented image, by using feature maps from the encoding stage, transpose convolution and concatenation operations.

This design allows the network to preserve both rich contextual information and fine details simultaneously. Moreover, U-Net2D incorporates skip connections between the layers of the encoding stage and their corresponding layers in the decoding stage.

These connections facilitate the direct propagation of high-level information throughout the network, aiding in the preservation and combination of features from different hierarchy levels. The U-Net2D architecture has been widely applied in various domains (Punn et al. [17]), and has proven to be a powerful and efficient choice for accurately segmenting structures in images.

1.3 Previous Work

Matamoros-Volante et al. [13] introduced a semi-automated analysis method that employed the Image-Based Focusing Criterion (IBFC) to investigate human sperm physiology.

The method utilized the IDEAS software (proprietary analysis software provided with the Amnis cytometry microscope) to detect focused images and generate a mask covering the head and flagellum, enabling statistical analysis and precise subcellular localization of labeling.

To identify the focused images, the method relied on the Gradient Root Mean Square (RMS), calculated by averaging the gradient normalized for intensity variations (Peli et al. [16]).

A threshold value of 62 was applied to the RMS to select focused images. However, a significant limitation of this methodology is its dependence on the threshold value, which can result in the incorrect inclusion of out-of-focus images.

Additionally, relying solely on the RMS value may not provide sufficient robustness, as out-of-focus images can still exhibit large RMS values.

Furthermore, the segmentation mask obtained through the IDEAS software may detect out-of-focus areas of the flagellum, which can potentially impact the biological results.

Recognizing the need for a more comprehensive approach to accurately detect focused images and segment the sperm head and flagellum, especially in sperm cells with diverse morphological features, our attention turned to the realm of deep learning methods.

In the realm of deep learning applied to sperm classification and segmentation, Fraczek et al. [8] proposed the utilization of Mask R-CNN for automated segmentation of the sperm head and flagellum from single-channel images. However, their results encountered difficulties in accurately segmenting the flagellum, achieving only a Mean Average Precision of 50% for the tail.

In another study by Movahed et al. [14], the segmentation problem was addressed through image pre-processing to enhance quality. They employed two independent CNN models optimized for head and flagellum segmentation, respectively.

Their experiments demonstrated that CNN-based models outperformed traditional segmentation algorithms such as SVM, Random Forest, Naive Bayer, Decision Trees, and KNN.

However, it's worth noting that their training dataset consisted of only 20 images, each containing many sperm.

Additionally, Marin et al. [12] compared the performance of U-Net and Mask R-CNN for head segmentation in a dataset comprising 19 images with a total of 210 sperm cells.

They evaluated two different categories for head segmentation: nucleus and acrosome. Their findings indicated that U-Net outperformed Mask R-CNN in this specific task, although the authors did not address the segmentation of the flagellum.

Regarding the classification of sperm images, to the best of our knowledge, there are no previous works on the use of deep learning for detecting focused images.

However, Riordon [18] proposed the use of the VGG16 network to identify different types of head shapes (normal, tapered, pyriform, and amorphous), achieving a true positive rate of 94%.

On the other hand, Spencer [20] employed an ensemble of deep learning networks (VGG16, VGG19, ResNet-34, and DenseNet-161) to classify the aforementioned head shapes, achieving F1-scores of 98.2 and 68.3 for the HuSHeM and SCIAN-MorphoSpermGS datasets, respectively.

For a comprehensive review of different techniques for classification and segmentation, refer to Suleman [21]. This research addresses the challenges of detecting focused images and accurately segmenting the sperm head and flagellum using deep learning.

Our solution involves two key steps. Firstly, we utilize a ResNet50 convolutional neural network to classify images as suitable for analysis (focused images), filtering out irrelevant ones.

Secondly, we employ a second CNN, U-Net 2D, to automatically segment the sperm structures (head and midpiece) in the selected images. This comprehensive approach enhances the efficiency and accuracy of sperm analysis.

2 Materials and Methods

Figure 1 provides a visual representation illustrating the main steps of the proposed methodology used for IBFC analysis. In the first step (Fig. 1 - Imaging), mouse sperm were obtained from the epididymis of 3 months old WT CD1 male mice according to Oliver et al. [15].

Images were acquired using flow-cytometry microscopy (Amnis® ImageStream®X Mk II imaging flow cytometer). In the second step (Fig. 1 - Dataset), the acquired images were categorized into two sets: focused images and out-of-focus images.

Additionally, a third set of images was created by manually segmenting the head and midpiece from the focused images. Finally, in the third step, the focused and out-of-focus images were used to train a ResNet 50 model, while the focused images with manual annotations were used to train a U-Net model.

This methodology enables the automatic analysis of IBFC images (Fig. 1 - Automatic Analysis). The trained ResNet model was applied to the image, and if the output suggests an out-of-focus image, the method rejects the analysis. Otherwise, if the image corresponds to a focused image, it is automatically segmented into midpiece and head regions using an U-Net model.

2.1 Imaging

Under the supervision of the Bioethics Committee at the Institute of Biotechnology, UNAM, spermatozoa were collected from the epididymis of 3-month-old WT male mice. The motile cell population was isolated using a swim-up separation technique and subsequently concentrated by centrifugation at 3000 rpm for 3 minutes.

Prior to image cytometry analysis, the cells were incubated on ice and stained with fluorescent probes: FM4-64 (500 nM) to label the membrane and Sytox blue (1 µM) to assess population viability. The suspended samples were prepared using a non-capacitating medium consisting of PBS (NaCl 137 mM, KCl 2.7 mM, Na2HPO4 10 mM, KH2PO4 1.8 mM).

Image acquisition was performed using the Amnis® ImageStream®X Mk II imaging flow cytometer, using the brightfield and fluorescence acquisition modalities, 405 nm and 561 nm lasers, a 60x objective, and a numerical aperture (NA) of 0.9. Images were captured over a period of 30 minutes, resulting in a total of over 100,000 images per experiment.

2.2 Manual Annotated Dataset

2.2.1 Dataset

Utilizing a methodology similar to that of Matamoros-Volante et al. [13], we employed the Root Mean Square (RMS) technique to generate two distinct sets of images: focused and out-of-focus.

This classification was based on the fluorescence signals of Sytox Blue and FM4-64 and may contain errors. To further refine the size of these image sets, we identified a subpopulation characterized by high FM4-64 fluorescence (indicating focused images) and low Sytox fluorescence (indicating out-of-focus images).

Finally, An experienced biologist manually selected a total of 297 focused images and 307 out-of-focus images from this refined selection process. Figure 2 visually illustrates examples of both focused and out-of-focus images.

2.2.2 Segmentation Dataset

We manually annotated the flagellum's midpiece and head of 294 focused images of mouse sperm in order to construct a ground-truth image dataset for segmentation.

The ImageJ software [1] was used to generate this dataset. For the midpiece annotation, we utilized the fluorescence channel and the "Segmented Line" option from the FIJI line selection tool. With a line width of 6, we carefully traced a line along the centerline of the flagellum, covering both its width and length.

This process resulted in a new image of the same size as the original, where the midpiece of the flagellum was clearly marked. Similarly, for the head masks, we used the brightfield channel and traced a line using the "Segmented Line" selection
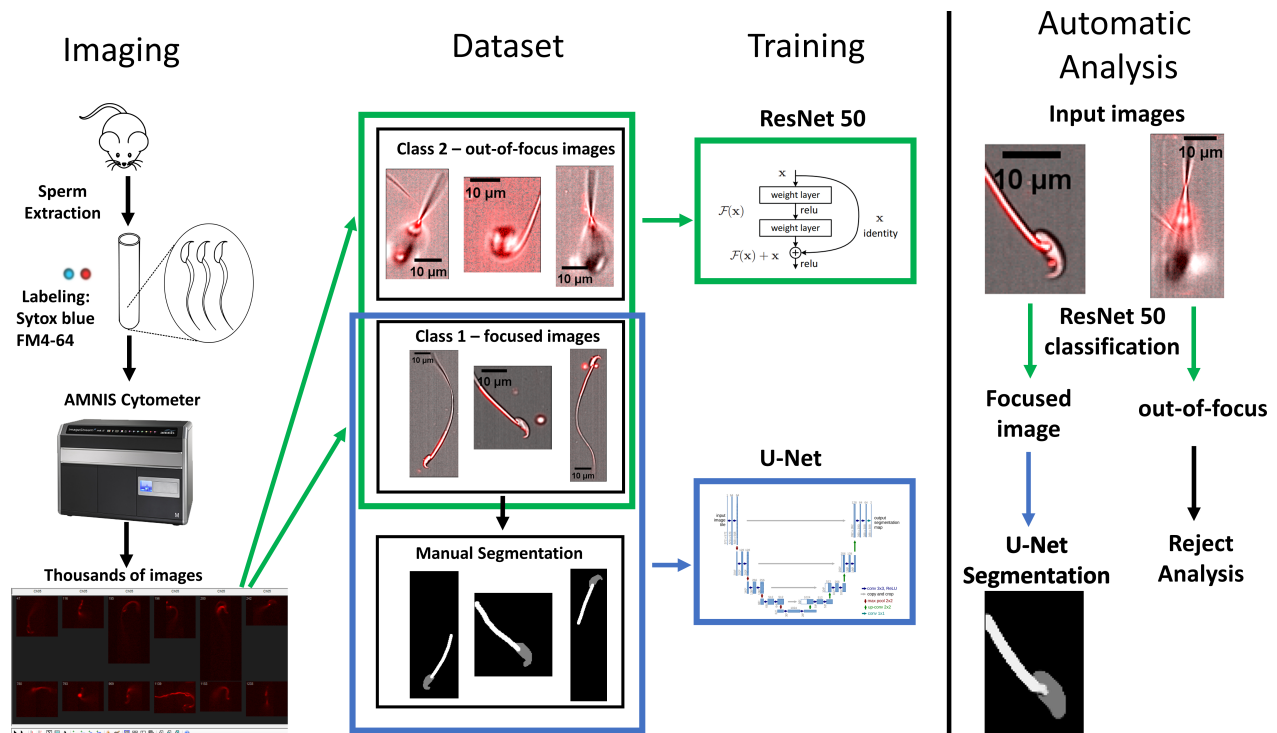
Fig. 1. Schematic representation of our proposed approach for IBFC analysis. Imaging: Sperm were obtained from the mouse epididymis and labeled using Sytox blue and FM4-64. Then, the AMNIS cytometer allowed us to obtain thousands of mouse sperm images. Dataset: Out-of-focus and focused images were manually selected from the thousands of images, along with manually annotated focused images. Training: This dataset was used to train ResNet50 for classification and the U-Net model for segmentation. Automatic Analysis: The trained models enable automatic IBFC analysis

tool with a width of one, accurately following the boundary of the head to create a closed region. The selection was then converted into a mask and filled using Fiji options. To provide a visual representation of the manually annotated images, please refer to Figure 3.

2.3 Image Pre-Processing for Deep Learning

Sperm flow cytometry images can have varying intensity values across different samples, occasionally containing outliers with extremely high or low intensities.

These outliers have the potential to impact the performance of deep learning algorithms. To address this, a preprocessing step was employed to ensure consistent and reliable intensity values.

In this step, we determine the low and high percentiles of intensity values. Specifically, we identify the values corresponding to the 1st and 99th percentiles.

Any intensity value below the low percentile is assigned the lowest value, while any value exceeding the high percentile is assigned the highest value. To further enhance comparability and consistency, the intensity values are then normalized to a standardized range of [0, 1].

In this normalization process, the low percentile value is mapped to 0, while the high percentile value is mapped to 1.

This ensures that all intensity values fall within the specified range, enabling easier comparison and analysis. Importantly, this preprocessing step was performed independently for each individual
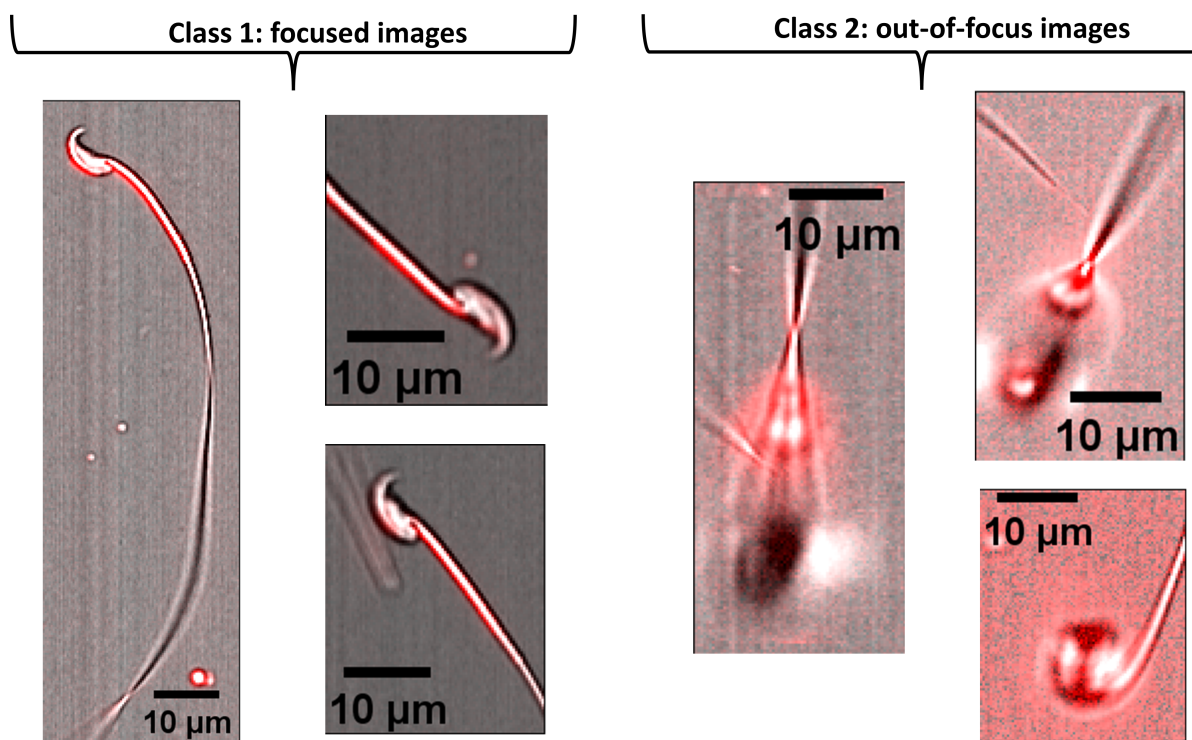
Fig. 2. A visual depiction of three samples representing focused images (2 channels, gray and red) from class 1, accompanied by three samples depicting out-of-focus images from class 2. The first channel (gray) represents bright-field, while the second channel (red) represents fluorescence

image and channel, allowing for the adaptability to variations within the dataset and ensuring accurate normalization of intensity values.

2.4 Training

2.4.1 Default Parameters

Training deep learning models (ResNet50 and U-Net 2D) was performed on a laptop equipped with an Nvidia GeForce GTX 1650 graphics card, which had 4 GB of memory available. The default parameters to train the networks were Adam optimizer with an initial learning rate of 0.001 and using betas set to (0.9, 0.999).

To ensure efficient training, a learning rate reduction strategy was implemented by decreasing the learning rate by a factor of 10 when there was no improvement in the loss function (reduction on plateau).

The loss function utilized was the sum of cross-entropy loss and dice loss. The dataset was randomly divided into training, validation, and test sets, with proportions of 80%, 10%, and 10% of the total dataset, respectively.

The batch size and number of epochs were the only parameters that varied depending on the specific task. It is important to note that all the models were trained from scratch. Finally, we utilized the PyTorch framework to train the deep learning models.

2.4.2 Classification Deep Learning Model

For the classification task, we utilized the classification dataset mentioned in the previous subsection, comprising 297 focused images and 307 out-of-focus images. Since the images in the dataset had different sizes, we standardized them by resizing each image to a fixed shape of
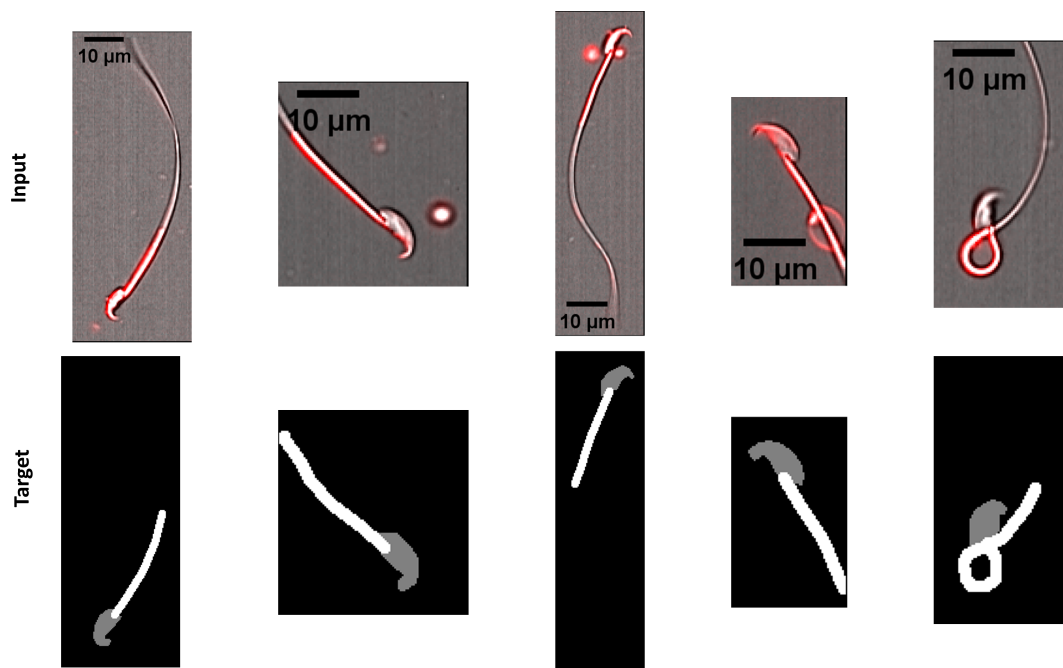
Fig. 3. Top row depicts input images, consisting of two channels. The first channel (red) represents bright-field, while the second channel (green) represents fluorescence. Bottom row depicts the corresponding target (ground-truth) image, also a two-channel image, with the first channel indicating head segmentation and the second channel representing the flagellum

(224, 224). To train the ResNet50 model, we used the default parameters, a batch size of 32, and conducted training for 50 epochs. The ResNet50 model was loaded from the torchvision library, the code used to train the network and predict images can be accessed from our GitHub repository at[1].

2.4.3 Segmentation Deep Learning Model

For the segmentation task, we utilized the segmentation dataset discussed in the previous subsection, which included 294 sperm images. The target images correspond to the manual annotations of the flagellum and head.

The U-Net model was trained with default parameters, a batch size of 8, and trained for 100 epochs. The dataset was divided into 236 training images, 29 validation images, and 29 test images.

To generate the U-Net model, custom functions were used, and the code for training the network and predicting images can be accessed from our GitHub repository at[2].

2.5 Evaluation Metrics

2.5.1 Classification Performance

To evaluate the performance of the ResNet50 network in a quantitative manner, we use the F1 score. The F1 score is calculated using the following formula:

$$\text{F1 score} = \frac{2\,\text{TP}}{2\,\text{TP} + \text{FP} + \text{FN}}. \tag{1}$$

In this formula, TP refers to the true positives, which are the number of images correctly predicted by the model as belonging to the class focused image.

---

[1]github.com/paul-hernandez-herrera/image_classification_pytorch

[2]github.com/paul-hernandez-herrera/unet_pytorch_2d

Table 1. Performance result of ResNet50 in 60 test images using 5 fold cross validation

#Run	0	1	2	3	4
F1 score	0.9915	0.9915	1.00	1.00	0.9830
Training time	7min 12s	7 min 13 s	7 min 17 s	7 min 13 s	7 min 14 s

FP represents the false positives, which are the number of images incorrectly predicted by the model as focused image class when the target class is actually out-of-focus; FN represents the false negatives, which are the number of images incorrectly predicted by the model as being out-of-focus when the target class is focused image.

The sum of FP and FN gives us the total number of errors made by the model, while TP represents the total number of correct predictions.

The F1 score is a metric that ranges between 0 and 1, where a value of 0 indicates the worst performance and a value of 1 indicates the best performance.

2.6 Segmentation Performance

To assess the performance of the U-Net model in segmentation tasks, we utilize the dice coefficient, which provides a qualitative measure. This metric quantifies the degree of overlap between two binary images, and it is calculated using the following formula:

$$\mathrm{Dice} = \frac{2|S \cap G|}{|S| + |G|}. \qquad (2)$$

In this formula, S represents the segmentation obtained from the U-Net model, and G represents the ground truth, which is the manually annotated head and flagellum.

The numerator, $|S \cap G|$, denotes the size of the intersection between the segmentation and ground truth. The denominator, $(|S| + |G|)$, is the sum of the sizes of the two images. Like the F1 score, the dice coefficient ranges between 0 and 1.

A value of 0 indicates no overlap between the ground truth and segmentation, representing the worst performance. Conversely, a value of 1 is a complete overlap, indicating that both the ground truth and segmentation are identical, which represents the best performance.

3 Results

In this section, we present the qualitative and quantitative results of our proposed methodology. To ensure more accurate results and robust findings, we performed 5-fold cross-validation. Each fold involved randomly selecting 80% of the images for training, 10% for validation, and 10% for testing purposes.

3.1 Classification

Our primary objective in this study was to identify the appropriate set of images for image analysis. Specifically, we aimed to distinguish between focused images suitable for analysis and out-of-focus images that are not.

To visually demonstrate this distinction, we provide examples in Figure 2. The figure showcases three instances of focused images alongside three instances of out-of-focus images. The focused images exhibit clear patterns, such as well-defined head shapes and bright intensity at the flagellum.

In contrast, these patterns are indiscernible in the out-of-focus images. Given the distinct visual differences between the two classes, our aim was to accurately separate these classes using a ResNet50 model.

To evaluate the performance of the ResNet50 model, we utilized a test set consisting of 60 images, which accounted for 10% of the classification dataset.

The evaluation metrics employed were the F1-score, which measures the model's accuracy, and the training time, which assesses the model's efficiency.

Table 1 provides a summary of the performance results for the 5 trained models corresponding to each fold. The obtained results showcase the remarkable capabilities of the ResNet50 model.

Table 2. Performance results $\mu(\sigma)$ of U-Net in 29 test images using 5-fold cross-validation. The average error (μ) and standard deviation (σ) are reported as evaluation metrics

#Run	0	1	2	3	4
Dice	0.8113 (0.09)	0.8144 (0.12)	0.8011 (0.14)	0.8001 (0.11)	0.8447 (0.07)
Training Time	32m 5s	32m 12s	32m 35s	41m 35s	39m 13s

With an average F1-score of 0.9932, the model's performance approached the maximum attainable F1-score, demonstrating its accuracy in distinguishing between focused and out-of-focus images, with a minor standard deviation of 0.0071, underscoring the consistent and reliable performance of the model.

Out of the 60 test images, the model only misclassified 1, 1, 0, 0, and 2 images, respectively, further affirming its efficacy. This result instills confidence in utilizing the ResNet50 model for detecting focused images in IBFC.

Additionally, the training time for each model was relatively quick, with an average of approximately 7 minutes and 14 seconds. This demonstrates the efficiency of training the ResNet50 network while still achieving commendable performance levels.

3.2 Segmentation: U-Net

Our second objective in this study was to automate the segmentation of the flagellum and the head within focused images.

To visually illustrate this objective, Figure 3 showcases representative images alongside their corresponding ground truth segmentation.

From the figure, it becomes evident that the expert biologist primarily focused on the green region of the image (fluorescence or channel 2).

Additionally, it can be observed that the flagellum exhibits varying shapes, including curved and rolled configurations.

However, a common characteristic among most flagellum shapes is their rectilinear structure. Similarly, the head appears regular and consistent across different sperm samples.

To effectively accomplish the automatic segmentation of these structures, we trained an U-Net architecture from scratch.

The performance of the U-Net model was evaluated using the dice similarity coefficient and training time. Table 2 presents a summary of the performance results for the 5 trained U-Net models corresponding to each fold.

Model 3 exhibited the lowest performance with an average dice score of 0.8001, while Model 4 achieved the best performance with an average dice score of 0.8447. The average dice score across all 5 models was 0.8141.

To provide a visual interpretation of the dice scores, Figure 4 illustrates examples of the results. In the first column, an image with a low dice score of 0.46 is displayed.

In this case, the U-Net model struggled to detect the sperm head, which is visually challenging due to its out-of-focus appearance. This suggests that such images should be excluded from the training set.

The second column of Figure 4 depicts an image with a dice score of 0.81, which is close to the average dice score of all models.

This image serves as a representative example of the average results achieved by the U-Net model. The output segmentation is visually like the ground-truth, with minor differences observed in the smoothness of the head segmentation.

The third column of Figure 4 showcases a test image with a high dice score of 0.92. Visually, it is difficult to distinguish any differences between the ground-truth and the model's output segmentation.

These results demonstrate that the U-Net model is capable of accurately segmenting the head and flagellum of the sperm. Additionally, the U-Net model exhibited fast training times, averaging at 35 minutes and 32 seconds.
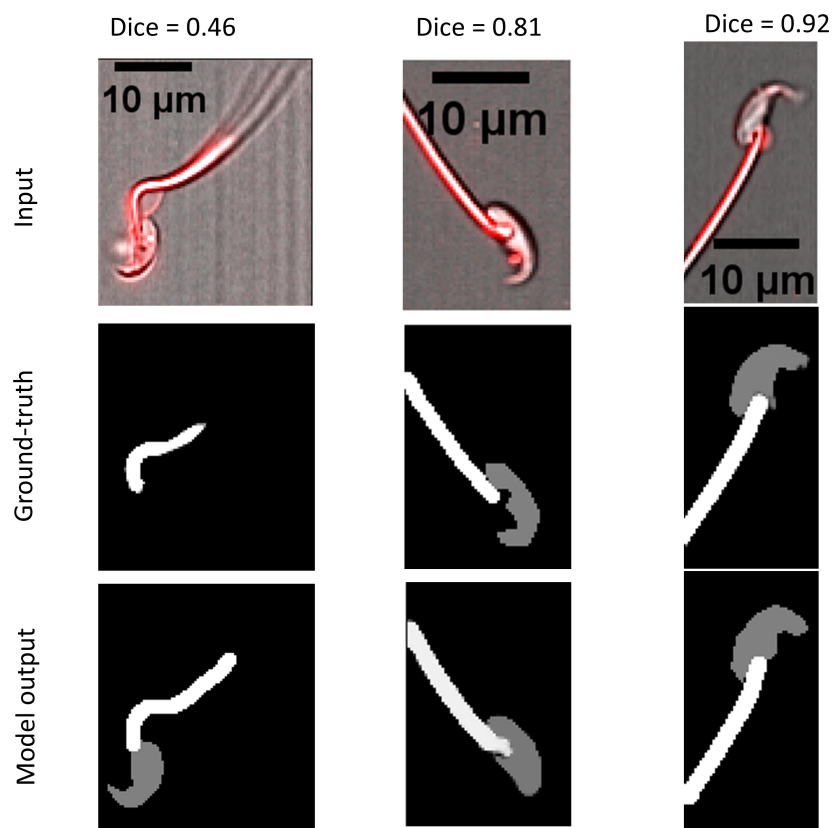
Fig. 4. A comparison of segmentation results using the dice coefficient, arranged in ascending order of values. Each row in the figure corresponds to a different test sample, with the columns representing the input image, ground-truth segmentation, and model output segmentation

4 Discussion

The most relevant previous work related to our approach is the one presented by Matamoros-Volante et al. [13]. There are several key differences between the two approaches.

Firstly, Matamoros-Volante's methodology employed human sperm, which has symmetrical head shape. In contrast, our approach is designed for mouse sperm, which have a more complex shape with a distinct hook-shaped head and a shorter flagellum.

Secondly, Matamoros-Volante's approach utilizes the segmentation mask obtained from the IDEAS software. However, this mask may include segments where the flagellum is out-of-focus or regions that are not relevant for the analysis.

For instance, their method employs a fixed dilation of 13 pixels to obtain the midpiece, without considering variations caused by rolling flagella or different sizes.

In contrast, our approach customizes the segmentation of the midpiece to the manual annotations, allowing it to adapt to the various sizes and shapes observed.

Finally, our approach offers the advantage of automatically detecting focused images, guided by the manual annotations of an expert biologist.

This automatic detection process is highly efficient. On the other hand, Matamoros-Volante's approach employs a semi-manual method (thresholding over the RMS value) to identify focused images, which may introduce errors during the detection process.

Considering all of these advantages, our deep learning-based approach to analyzing IBFC images has the potential to greatly improve and innovate the way these images are analyzed.

5 Conclusions

IBFC plays a crucial role in the acquisition of a large number of images. However, manual analysis of these images is impractical. In this study, we propose an innovative approach that leverages convolutional neural networks (CNNs) to address this challenge effectively.

Our approach enables the identification and exclusion of uninformative images, while also automating the segmentation of the flagellum's midpiece and head using the U-Net architecture.

The quantitative results obtained from the ResNet50 classification exhibit exceptional performance, achieving an average F1 score of 0.99. This signifies near-perfect classification accuracy. On the other hand, the U-Net segmentation demonstrates decent performance, with an average dice coefficient of 0.81.

The qualitative results reveal a high level of similarity between the U-Net output and the ground truth, indicating the effectiveness of the segmentation approach. Moreover, we demonstrated that the deep learning approach can be trained using a laptop, resulting in fast training times without the need for expensive equipment.

By combining CNN-based classification and segmentation techniques, our approach presents a compelling solution for the automated analysis of flow cytometry images. It offers significant potential for accelerating research, improving efficiency, and enabling more comprehensive biological investigations.

In future work, we plan to optimize the parameters of the U-Net model further by experimenting with different loss functions. Additionally, we aim to increase the size of the training set and apply data augmentation techniques to enhance the model's performance.

Although our results visually appear very similar to the ground truth, we have not yet compared them with the methodology presented by Matamoros-Volante et al. [13].

Therefore, we will conduct comparisons between the two approaches for IBFC analysis. Moreover, we intend to apply this methodology to extract valuable biological insights and draw conclusions from comparing sperm in different functional states.

Acknowledgments

References

1. **Abràmoff, M. D., Magalhães, P. J., Ram, S. J. (2004).** Image processing with imagej. Biophotonics international, Vol. 11, No. 7, pp. 36–42.

2. **Adan, A., Alizada, G., Kiraz, Y., Baran, Y., Nalbant, A. (2017).** Flow cytometry: Basic principles and applications. Critical Reviews in Biotechnology, Vol. 37, No. 2, pp. 163–176. DOI: 10.3109/07388551.2015.1128876.

3. **Alnuaim, A. A., Zakariah, M., Shashidhar, C., Hatamleh, W. A., Tarazi, H., Shukla, P. K., Ratna, R. (2022).** Speaker gender recognition based on deep neural networks and ResNet50. Wireless Communications and Mobile Computing, Vol. 2022, pp. 1–13. DOI: 10.1155/2022/4444388.

4. **Anand, V., Gupta, S., Nayak, S. R., Koundal, D., Prakash, D., Verma, K. D. (2022).** An automated deep learning models for classification of skin disease using dermoscopy images: A comprehensive study. Multimedia Tools and Applications,

Vol. 81, No. 26, pp. 37379–37401. DOI: 10.1007/s11042-021-11628-y.

5. **Barteneva, N. S., Fasler-Kan, E., Vorobjev, I. A. (2012).** Imaging flow cytometry: Coping with heterogeneity in biological systems. Journal of Histochemistry & Cytochemistry, Vol. 60, No. 10, pp. 723–733. DOI: 10.1369/0022155412453052.

6. **Basiji, D. A., Ortyn, W. E., Liang, L., Venkatachalam, V., Morrissey, P. (2007).** Cellular image analysis and imaging by flow cytometry. Clinics in Laboratory Medicine, Vol. 27, No. 3, pp. 653–670. DOI: 10.1016/j.cll.2007.05.008.

7. **Doan, M., Vorobjev, I., Rees, P., Filby, A., Wolkenhauer, O., Goldfeld, A. E., Lieberman, J., Barteneva, N., Carpenter, A. E., Hennig, H. (2018).** Diagnostic potential of imaging flow cytometry. Trends in Biotechnology, Vol. 36, No. 7, pp. 649–652. DOI: 10.1016/j.tibtech.2017.12.008.

8. **Fraczek, A., Karwowska, G., Miler, M., Lis, J., Jezierska, A., Mazur-Milecka, M. (2022).** Sperm segmentation and abnormalities detection during the intracytoplasmic sperm injection procedure using machine learning algorithms. 15th International Conference on Human System Interaction, pp. 1–6. DOI: 10.1109/hsi55341.2022.9869511.

9. **He, K., Zhang, X., Ren, S., Sun, J. (2016).** Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. DOI: 10.1109/cvpr.2016.90.

10. **Hossain, M. B., Iqbal, S. H. S., Islam, M. M., Akhtar, M. N., Sarker, I. H. (2022).** Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest x-ray images. Informatics in Medicine Unlocked, Vol. 30, pp. 100916. DOI: 10.1016/j.imu.2022.100916.

11. **Jiang, H., Diao, Z., Shi, T., Zhou, Y., Wang, F., Hu, W., Zhu, X., Luo, S., Tong, G., Yao, Y. D. (2023).** A review of deep learning-based multiple-lesion recognition from medical images: Classification, detection and segmentation. Computers in Biology and Medicine, pp. 106726. DOI: 10.1016/j.compbiomed.2023.106726.

12. **Marín, R., Chang, V. (2021).** Impact of transfer learning for human sperm segmentation using deep learning. Computers in Biology and Medicine, Vol. 136, pp. 104687. DOI: 10.1016/j.compbiomed.2021.104687.

13. **Matamoros-Volante, A., Moreno-Irusta, A., Torres-Rodriguez, P., Giojalas, L., Gervasi, M. G., Visconti, P. E., Treviño, C. L. (2018).** Semi-automatized segmentation method using image-based flow cytometry to study sperm physiology: the case of capacitation-induced tyrosine phosphorylation. MHR: Basic science of reproductive medicine, Vol. 24, No. 2, pp. 64–73. DOI: 10.1093/molehr/gax062.

14. **Movahed, R. A., Mohammadi, E., Orooji, M. (2019).** Automatic segmentation of sperm's parts in microscopic images of human semen smears using concatenated learning approaches. Computers in Biology and Medicine, Vol. 109, pp. 242–253. DOI: 10.1016/j.compbiomed.2019.04.032.

15. **Oliver, E. I., Jabloñski, M., Buffone, M. G., Darszon, A. (2023).** Two-pore channel 1 and ca2+ release-activated ca2+ channels contribute to the acrosomal ph-dependent intracellular ca2+ increase in mouse sperm. The Journal of Physiology. DOI: 10.1113/jp284247.

16. **Peli, E. (1990).** Contrast in complex images. JOSA A, Vol. 7, No. 10, pp. 2032–2040. DOI: 10.1364/josaa.7.002032.

17. **Punn, N. S., Agarwal, S. (2022).** Modality specific U-Net variants for biomedical image segmentation: A survey. Artificial Intelligence Review, Vol. 55, No. 7, pp. 5845–5889. DOI: 10.1007/s10462-022-10152-1.

18. **Riordon, J., McCallum, C., Sinton, D. (2019).** Deep learning for the classification of human sperm. Computers in biology and medicine, Vol. 111, pp. 103342. DOI: 10.1016/j.compbiomed.2019.103342.

19. Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241. DOI: 10.48550/ARXIV.1505.04597.

20. Spencer, L., Fernando, J., Akbaridoust, F., Ackermann, K., Nosrati, R. (2022). Ensembled deep learning for the classification of human sperm head morphology. Advanced Intelligent Systems, Vol. 4, No. 10, pp. 2200111. DOI: 10.1002/aisy.202200111.

21. Suleman, M., Ilyas, M., Lali, M. I. U., Rauf, H. T., Kadry, S. (2023). A review of different deep learning techniques for sperm fertility prediction. AIMS Mathematics, Vol. 8, No. 7, pp. 16360–16416. DOI: 10.3934/math.2023838.

22. Zhang, F., Lei, C., Huang, C. J., Kobayashi, H., Sun, C. W., Goda, K. (2019). Intelligent image de-blurring for imaging flow cytometry. Cytometry Part A, Vol. 95, No. 5, pp. 549–554. DOI: 10.1002/cyto.a.23771.

23. Zou, J., Han, Y., So, S. S. (2009). Overview of artificial neural networks. Artificial neural networks: Methods and applications, pp. 14–22. DOI: 10.1007/978-1-60327-101-1_2.