cytoKnet: An instance segmentation approach for multiple myeloma plasma cells using conditional kernels

Received: 18 July 2023 | Revised: 9 September 2023 | Accepted: 11 October 2023

DOI: 10.1002/ima.22984

RESEARCH ARTICLE

WILEY

cyto-Knet: An instance segmentation approach for multiple myeloma plasma cells using conditional kernels

Massimo Salvi | **Nicola Michielli** | **Kristen M. Meiburger** | **Cristina Cattelino** | **Bruna Cotrufo** | **Matteo Giacosa** | **Chiara Giovanzana** | **Filippo Molinari**

Biolab, PoliToBIOMed Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy

Correspondence
Massimo Salvi, Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin 10129, Italy.
Email: massimo.salvi@polito.it

Abstract

Multiple myeloma disrupts normal blood cell production, requiring early detection due to the increased risk of bone metastases. Although various artificial intelligence (AI) methods have been developed to assist pathologists, they often lack comprehensive metrics to measure both detection and segmentation errors. This study presents a hybrid framework that combines deep learning and heuristic techniques to achieve accurate instance segmentation of individual plasma cells in Giemsa-stained cytology images. Our proposed network, called cyto-Knet, incorporates an innovative color-balancing method as a preprocessing step to standardize the appearance of cytological images. Our network leverages a 4-class segmentation strategy with conditional kernels to enhance segmentation performance and accuracy. Additionally, a marker-based watershed algorithm is employed in the postprocessing step to address the challenge of merged objects. Extensive validation at both pixel and object-based levels demonstrates superior performance compared with state-of-the-art techniques. Our method achieves pixel-based metrics (precision, recall, and F1-score) of approximately 0.90. The object-based evaluation reveals an average Dice coefficient of 0.9130 and an aggregated Jaccard index of 0.8237. Importantly, our solution is designed for integration into an end-to-end system for diagnosis support and can be easily extended to other applications.

KEYWORDS

cytology, deep learning, Giemsa stain, instance segmentation, myeloma plasma cells

1 | INTRODUCTION

Multiple myeloma (MM) is a cancer that disrupts the normal production of blood cells by causing the proliferation and accumulation of plasma cells in the bone marrow. This type of malignancy accounts for 10% of all blood cancer cases and primarily affects men over the age of 65.[1,2] The median survival rate for MM patients is typically 5–7 years and varies based on factors such as tumor stage, cytogenetic abnormalities, and response to treatment.[3] Early detection is crucial as MM often leads to bone metastases.

Currently, the gold standard for MM diagnosis involves visually examining a serum sample obtained through biopsy to identify abnormal plasma cells. These cells exhibit peculiar characteristics such as irregular shape, abnormal nuclei, and enlarged size. Pathologists analyze the sample and determine the percentage of abnormal plasma cells present, known as the plasma cell percentage, which is a critical factor in MM diagnosis. However, this manual process can be time-consuming and prone to interoperator variability. To address these limitations, automated algorithms have been developed to enable faster and more consistent segmentation of malignant cells.[4,5]

The challenge of segmenting cells, including identifying nuclei and cytoplasm, can be addressed using instance segmentation techniques, which combine pixel-wise prediction and object detection.[6] In digital cytology, instance segmentation of plasma cells aims to differentiate individual or clustered cells with different sizes and shapes. Furthermore, it must handle potential issues arising from suboptimal staining processes that may affect the chromatic appearance of cellular structures. The main challenges associated with cytological cell segmentation are illustrated in Figure 1.

Recent studies in the field of instance segmentation have focused on using deep networks, specifically convolutional neural networks (CNNs), to identify different cellular components of interest in the microscopic image, including nucleus and cytoplasm. The U-Net architecture is commonly combined with other networks for this purpose. For instance, Bozorgpour et al.[7] applied U-Net to segment each nucleus instance in the input images, while the Attention Deeplabv3+ model generated the final segmentation map. Azad et al.[8] employed the deep frequency re-calibration U-Net to segment both nuclei and cytoplasm. To address instance segmentation, Pandey et al.[9] fused transformer-based models with CNN backbones. Faura et al.[10] utilized three different networks—Mask R-CNN, Hybrid Task Cascade (HTC), and Sample Consistency Network (SCNet) —to automatically segment plasma cells in MM patients. They employed ResNet and the recent ResNeSt convolutional backbones as feature extractors. Another recent study proposed DCSAU-Net,[11] a deeper and more compact split-attention U-Net for medical image segmentation, which was compared against a wide range of state-of-the-art models.

The use of neural network-based algorithms has demonstrated significant potential in improving the speed and reproducibility of MM diagnosis. However, a literature gap exists in this application as the majority of the implementation and training of black-box models are without a pre- and postprocessing strategy to ensure the stability and robustness of the entire end-to-end pipeline. Additionally, the data acquisition setup may negatively impact the training phase due to high color variation observed in images captured using different devices like microscopes or cameras. The use of resource-intensive ensemble models significantly increases computational times, and limited training data for such data-intense approaches may necessitate the adoption of massive data augmentation strategies, risking overfitting. Despite the progress in deploying increasingly complex architectures, the evaluation of these methods still lacks on quantitative metrics that effectively address object detection and segmentation errors in a unified manner.[12] False-positive instances have not been adequately assessed, which can have a negative impact on the clinical perspective of plasma cell counting. Furthermore, the instance segmentation task presents inherent complexities, such as (i) suboptimal image acquisition and quality, (ii) partially overlapping cells, and (iii) poorly defined object contours.

The aim of this research paper is to develop a hybrid framework that combines recent deep learning methods with heuristic techniques to achieve precise and accurate segmentation of individual plasma cells, which is a crucial requirement for this application. We used the dataset provided by the SegPC-2021 challenge,[13] to develop our method. In this challenge, microscopic images were

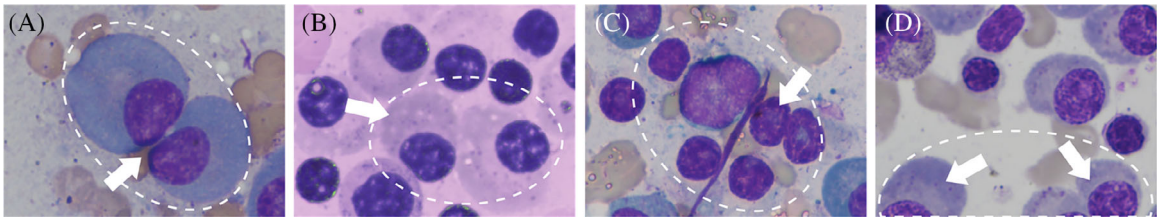

FIGURE 1 Challenges in the dataset for instance segmentation of cytological images. The instance segmentation task for cytological images presents several challenges, including (A) clustered plasma cells with merged boundaries, where the cytoplasm of two cells may touch each other, the cytoplasm of one cell and the nucleus of another may touch each other, or cell nuclei may touch each other; (B) cytoplasm stain color may be similar to the background color, making the cell boundary indistinguishable; (C) unwanted stained cells may be annotated due to the subjective clinical operator's experience; and (D) partially captured objects may be present in the images.

obtained from bone marrow aspirate slides of patients diagnosed with MM. The main contributions of this paper can be summarized as follows:

- We propose a novel color-balancing method that effectively standardizes the appearance of cytological images acquired from different devices, such as microscopes and cameras. This innovative approach addresses variations in illumination conditions by normalizing background color and calibrating staining within images.
- We introduce a new image segmentation network called cyto-Knet, which combines vision transformers and multiscale feature extraction. An important aspect of our approach is the utilization of a four-class training strategy that includes an additional cell border class. Furthermore, we employ conditional kernels for each class, resulting in optimized segmentation accuracy.
- We present a robust framework that incorporates a highly effective postprocessing phase designed to enhance the performance of our deep learning model. To address the issue of merged objects and reduce false-positive cell segmentations, we employ the watershed algorithm and a tailored postprocessing step.
- We conduct extensive validation using both pixel and object-level metrics to fully assess errors and accuracy. This approach is essential as it addresses a current gap in existing methods that often overlook segmentation errors. We compare our results with prior published techniques and also conduct an ablation study to demonstrate the contributions of our pre- and postprocessing components.

The rest of this paper is organized as follows: Section 2 provides an in-depth description of the proposed method; experimental results are reported and discussed in Sections 3 and 4, respectively. Finally, Section 5 includes a final summary and future lines of research.

2 | MATERIALS AND METHODS

In this paper, we present a fully supervised method for myeloma cell segmentation in cytological images. Figure 2 summarizes the overall flowchart of our approach. Our method follows a three-step procedure: (i) preprocessing and patch extraction, (ii) cell segmentation using conditional kernels, and (iii) postprocessing and segmentation refinement. A detailed description of our approach is provided in the following subsections.

2.1 | Dataset

We used the dataset from the SegPC-2021 challenge[13] to develop our method. This challenge involved acquiring microscopic images from bone marrow aspirate slides of patients diagnosed with MM. The slides were stained using the Jenner-Giemsa stain, and the plasma cells were manually segmented by experts and provided as part of the challenge dataset. The images were digitized in BMP file format using two different cameras attached to the microscope. The Olympus CellSens software (version 2.1) was used to capture 690 images with a size of 1536×2040 pixels, and the Nikon DS-5 M camera acquired 85 images with a size of 1920×2560 pixels.[14–16]

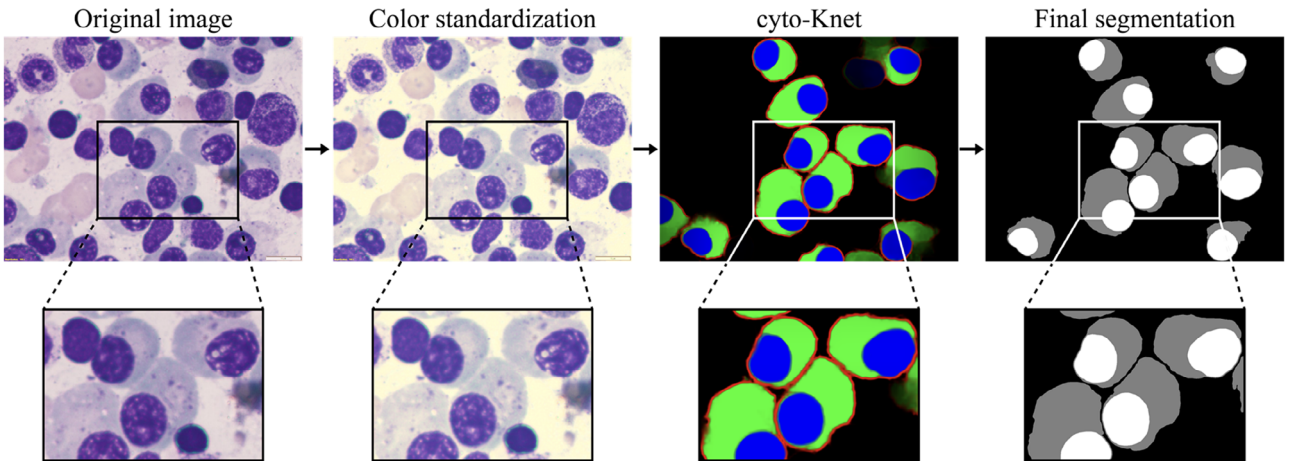

FIGURE 2 Schematic representation of our segmentation pipeline with a zoomed-in view of the result for each step. The process begins with the application of color standardization to balance the chromatic components of the original cytological image. Subsequently, our deep neural network, called cyto-Knet, is employed for the segmentation task. Finally, an ad-hoc postprocessing technique is applied to accomplish the instance segmentation of nuclei and cytoplasm for each plasma cell.

Each image contained multiple plasma cells, and each cell was segmented into three classes: nuclei, cytoplasm, and background.

The entire dataset consisted of 775 microscopic images, which were divided into a training set with 298 samples, a validation set with 200 images, and a test set with 277 samples. Manual annotations were provided only for the training and validation sets.[13] Since the test data from the challenge are not publicly available yet, we randomly divided the 498 images into a construction set, with 400 images used for training and 48 images for validation. The remaining 50 images were reserved as a blind test set.

2.2 | cyto-Knet preprocessing: novel color balancing and patch extraction

Preprocessing is a crucial step in enhancing the contrast between cells and background. However, conventional contrast stretching methods may compromise the integrity of the image, introducing potential biases during analysis.[17] For this reason, we have developed a novel preprocessing technique based on color balancing. This technique standardizes the conditions of image acquisition while preserving the image's information content, reducing chromatic variability, and achieving a more natural white background appearance. Our color-balancing method comprises two steps: color calibration and illuminant correction.

In the color calibration phase, the image is converted to the LAB color space. Then, the A and B layers are adjusted according to Equation 1 and Equation 2, respectively. The modified image is then converted back to the RGB color space using Equation 3.

$$A_{new} = A - 1.1 * (avgA - 128) * {}^{L}\!/_{255} \quad (1)$$

$$B_{new} = B - 1.1 * (avgB - 128) * {}^{L}\!/_{255} \quad (2)$$

$$IMG_{new} = lab2rgb(L, A_{new}, B_{new}) \quad (3)$$

In these equations, *avgA* and *avgB* represent the average values of A-layer and B-layer in the initial LAB image. Equation 1 adjusts the A channel relative to the L channel proportionally to the difference between *avgA* and 128. This quantity is then scaled by a factor of 1.1. Through visual comparison, we found factors less than 1 inadequate to fully correct potential image calibration errors, while factors slightly above 1 were effective. Equation 2 performs the same operations on the B channel. These adjustments aim to eliminate background polarization caused by various image acquisition conditions like ambient lighting and uncalibrated camera colors, as shown in Figure 3.

In the illuminant correction step, we employ a white balance method to standardize all image colors based on the color of the white background.[18,19] To correct for illumination variations in the image, we first identify the mask of bright pixels by applying a threshold in the L-channel of the LAB color space. Then, we calculate the mean value in grayscale (*gv*) for each channel of the image. This parameter represents the intensity of the bright pixels in the uncorrected image. To adjust the illumination, we define a value called *wv*, which represents the desired intensity of the bright pixels in the calibrated image. For this study, we used a fixed value of 0.80 for all channels. We chose this value to avoid a background that is too light ($wv = 1.0$) or too dark ($wv = 0.60$). If the value of *wv* is higher than *gv*, the brightness of the image is increased; otherwise, it is decreased. Next, each channel of the RGB image is converted into optical density (OD) and normalized to the *gv* value using Equation 4. Finally, to return to the RGB space, we apply the inverse formula to multiply the previous value by *wv* using Equation 5.

$$OD^{i}_{NORM} = -\log_{10}\left(IMG^{i}_{new}/gv^{i}\right), \forall i = 1,2,3 \quad (4)$$

$$C^{i}_{NORM} = wv \cdot 10^{-OD^{i}_{NORM}}, \forall i = 1,2,3 \quad (5)$$

This approach effectively adjusts the brightness of the image, ensuring that no areas are over- or underexposed. Figure 3 provides an example of the effectiveness of the preprocessing technique. Our method accurately standardizes different color presets within the image, while simultaneously preserving the image's information content and reducing chromatic variability. After performing image standardization, we extracted patches of 1500 × 1500 pixels with a spatial overlap of 75% to fit the network input size.

2.3 | cyto-Knet core: cell instance segmentation using conditional kernels

When considering the task of semantic segmentation of plasma cells, it is essential for the final framework to utilize information from different scales of the image.[13] This task presents two primary challenges: firstly, nuclei often merge and form clusters, and secondly, the cytoplasm of multiple cells can come into contact with each other. To overcome these challenges, we have developed a novel architecture called cyto-Knet, as depicted in Figure 4. This architecture incorporates a pyramid pooling module

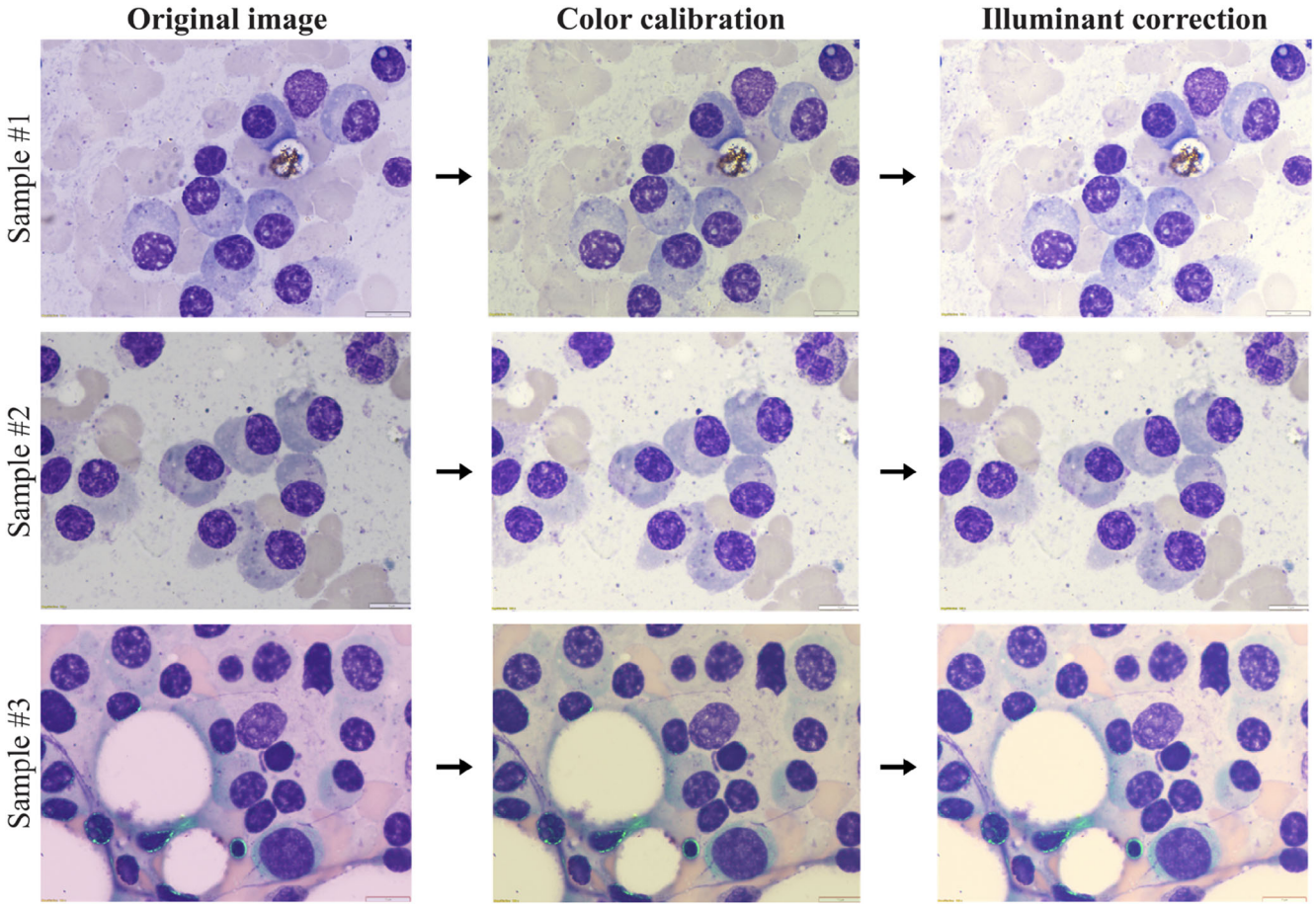

FIGURE 3 Preprocessing step applied to the cytological image. As can be seen, this standardization approach calibrates images effectively and consistently, reducing chromatic variability and ensuring that the image's information content is preserved.

to extract multiscale features and performs precise segmentation by using conditional kernels customized for each class.

The cyto-Knet model integrates the UPerNet (Unified Perceptual Parsing for Scene Understanding)[20] into the K-Net architecture.[21] Additionally, this model employs a Swin transformer[22] as a feature extractor, which provides a robust and accurate representation of the input data. The Swin transformer constructs hierarchical feature maps by merging image patches in deeper layers. One of the advantages of this architecture is its linear computational complexity in relation to the input image size, achieved by conducting self-attention computations solely within local windows, as indicated by the red boxes in Figure 4A. Following feature extraction, the UPerNet leverages global context information by merging features into different pyramid scales. This architecture combines a bottom-up pathway, which captures low-level features of the input image, with a top-down pathway that refines the segmentation map using skip connections (Figure 4B). The lateral connection between the bottom-up and top-down pathways merges feature maps of different resolutions to further enhance the segmentation results. Finally, the K-Net module updates conditional kernels based on their corresponding segmentation class, resulting in a more accurate mask prediction.

The conditional kernels in cyto-Knet are a type of convolutional kernel that dynamically adjusts its behavior based on the input data. Unlike traditional convolutional kernels that apply the same weights to all input patches, conditional kernels can adapt their weights based on the characteristics of the input. To achieve this, cyto-Knet employs a gating mechanism that determines the contribution strength of each conditional kernel. The gating mechanism takes the input data and produces a set of gating values, which determine the strength of each kernel's contribution. Kernel refinement involves updating each kernel with contextual information from other groups through a multihead attention block. Convolutions between the obtained kernels and the feature map result in a more precise mask prediction, improving the discriminative ability of the kernels and enhancing the final segmentation performance (Figure 4C).

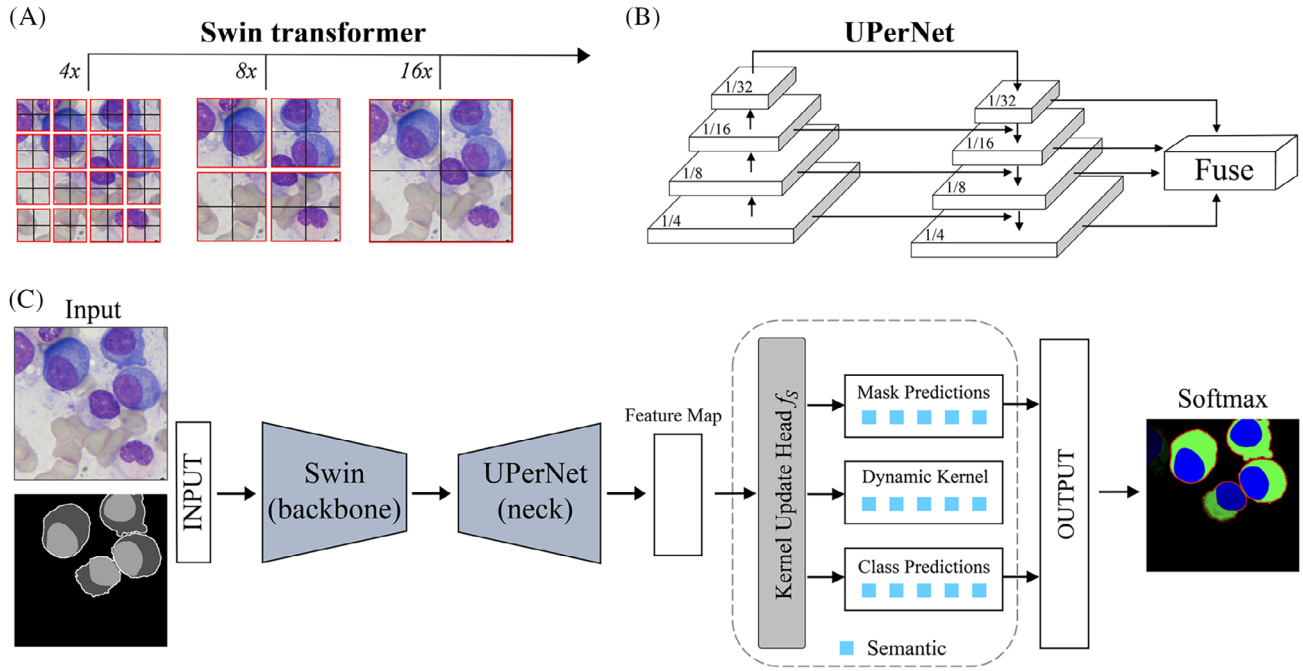

FIGURE 4 Segmentation network used in this work. (A) Features are extracted by using a Swin transformer as backbone. (B) Multiresolution features aggregation is performed by the UPerNet (Unified Perceptual Parsing for Scene Understanding). (C) Overall architecture of the cyto-Knet. During the inference process, the network generates a softmax output that indicates the probability of each pixel belonging to one of the following classes: background (shown in black), nucleus (blue), cytoplasm (green), and cell border (red).

Another novel aspect of this work is the training strategy in terms of the number of classes used to train the network. Specifically, the network is trained on a four-class problem using 1500 × 1500 RGB images as input and labeled masks as the target. The pixels are labeled in four classes: nuclei, cytoplasm, background, and the newly introduced class for the cell boundary, an approach that has not yet been explored in this dataset. The inclusion of both the object and the border as classes serves to accurately define the spatial boundaries of each cell by leveraging information on the location and contour of each object, as demonstrated in Figure 4. The use of a fourth class to represent the cell boundary is crucial for the postprocessing stage to effectively separate individual cells (Section 2.4). To determine the cell boundary class, which was not originally provided with the challenge dataset, the outline of each plasma cell is extracted from the manual mask, and a dilation is performed using a 5-pixel radius disk to obtain a binary mask of cell contours. During training, real-time data augmentation is implemented by randomly flipping and rotating both the input image and the corresponding encoded mask to increase the diversity of the training data.

The cyto-Knet was trained on 1216 patches with a minibatch size of 8, using a Dice loss and AdamW optimizer for both the decode and auxiliary head of the model. The maximum number of epochs was set to 100 with a validation patience of 5 epochs for early stopping. The model was trained for 7 h on a dedicated workstation equipped with a GeForce RTX3090, a 4.5 GHz 10-core CPU, and 128 GB of RAM.

2.4 | cyto-Knet postprocessing: patch aggregation and segmentation refinement

The proposed segmentation model works on patches of size 1500 × 1500 pixels. However, the images in the SegPC challenge have different sizes depending on the acquisition device, either 1920 × 2560 or 1536 × 2040 pixels. To address this issue, we employed the same patch aggregation strategy adopted in our previous works.[23,24] This technique involves creating an extended image by padding the original image with mirror reflections. A sliding window operator with overlap is then applied, and each window is processed by the network. The final softmax is obtained by center cropping the resulting heatmap, which represents the pixel-wise probability for the "nucleus," "cytoplasm," "border," or "background" classes.

The raw segmentation mask, illustrated in Figure 5, is generated through majority voting from the 4-class softmax probability map mentioned earlier. Subsequently, a postprocessing step is conducted to refine the

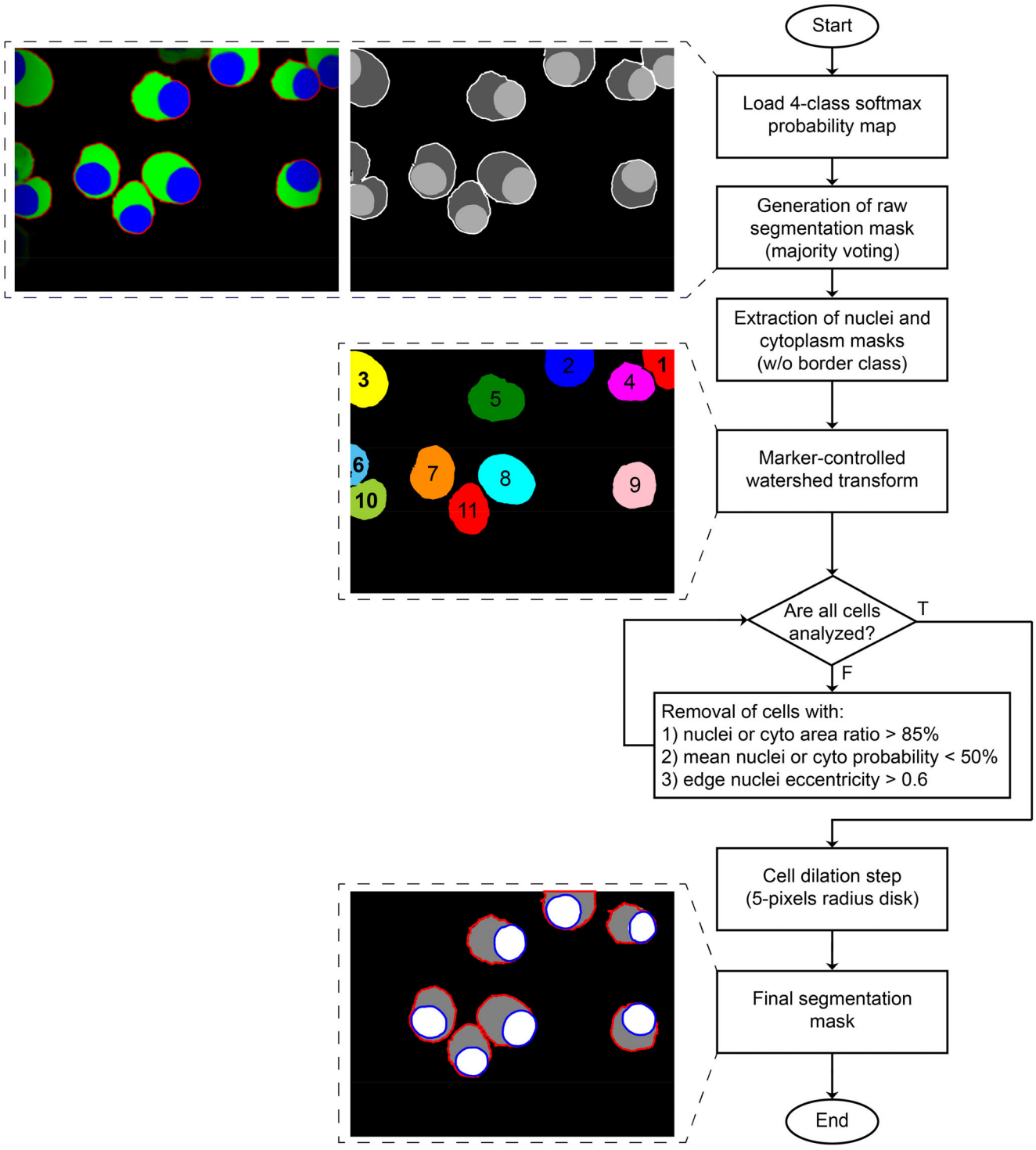

FIGURE 5 Graphical flowchart representation of the postprocessing strategy adopted in this work.

segmentation of individual plasma cells, especially in the presence of clustered objects. Nuclei and cytoplasm masks are extracted from the raw segmentation by merging the cell border class with the background, facilitating the separation of cells. This merging process reduces the number of clustered cells and results in a clearer division between single cells. Additionally, image cleaning operations such as small objects removal and hole filling are applied to the segmentation masks.

The marker-controlled watershed algorithm based on the Euclidean distance transform[25,26] is then applied to separate plasma cells into single instances. Markers are defined as local maxima of the distance from the background class. Each instance is individually processed and

analyzed using a heuristic rule-based suppression step. In cases where multiple nuclei exist for the same cell, only the nucleus with the largest area is retained. Instances consisting solely of a nucleus or cytoplasm are removed if their area-to-cell ratio exceeds 85%, as they do not represent the cells of interest. Cells with an average softmax probability of the nucleus or cytoplasm lower than 50% are also discarded due to potential network misclassification. Furthermore, cells with nuclei located on the image's edge and an eccentricity higher than 0.6 (indicating a shape very different from a circle) are considered partially acquired objects and are excluded.

The final segmentation mask is obtained by applying a dilation operation with a 5-pixel radius disk to retrieve the newly introduced border class. Figure 5 provides a systematic diagram illustrating the key steps of the postprocessing strategy. In the 4-class softmax probability map, the nuclei, cytoplasm, border, and background classes are represented by the colors blue, green, red, and black, respectively. The final segmentation mask, obtained after postprocessing, is displayed with overlapping ground truth nuclei and cell contours shown in blue and red, respectively.

2.5 | Performance metrics

The SegPC-2021 challenge introduced the mean Intersection over Union (mIoU) score as an evaluation metric.[13] This score measures the overlap between each ground truth object and detected instance using the Jaccard index. If no automated segmentation is found, indicating a false-negative object, a score of zero is assigned. The mIoU is calculated as the average value across all ground truth instances. However, the mIoU does not consider unclaimed segmented instances (false-positive objects), which is a limitation of this metric. Additionally, the images are resized to 1080×1440 pixels before computing the mIoU score to standardize image sizes. The score is then calculated as the average over all plasma cells in each subset, rather than over images as with other quantitative metrics.

To address these limitations, we compute several pixel and object-based metrics to provide a more robust performance comparison. At the pixel level, we calculate various metrics by comparing the manual and automatic segmentation masks, including precision, recall, and the F1 score (also known as Dice coefficient). Furthermore, our work introduces object-level metrics to evaluate individual instances using the same metrics but with an object-wise definition.[27] Additionally, we compute the Hausdorff distance (HD) to measure object contour matching.[28] For a more comprehensive assessment, we calculate the aggregated Jaccard index (AJI), which effectively penalizes both pixel- and object-level segmentation and detection errors.[12] These metrics collectively provide a thorough and robust evaluation of our segmentation model's performance.

3 | RESULTS

3.1 | Hyperparameter tuning

In this study, we perform a fine-tuning of the cyto-Knet architecture to identify its most effective components. We evaluate various backbone networks, including ResNet, DenseNet, and the Swin transformer, with different depths and numbers of convolutional blocks. Based on the analysis, we select the Swin transformer as it achieves the best results on the validation set. We also experiment with different hyperparameters such as learning rate, batch size, loss function, and weight for the decode and auxiliary head of the network (Table 1). The performance of each configuration is assessed on the validation set, and the final architecture with the highest Dice score is chosen.

3.2 | Plasma cell segmentation

The pixel- and object-based metrics for the segmentation of nuclei, cytoplasm, and the whole plasma cell, are reported in Table 2. For cytoplasm segmentation, only pixel-based metrics can be computed due to instances referring to multiple connected components with nonconvex shapes. The proposed cyto-Knet achieves an mIoU of 0.8561 for the construction set and an mIoU of 0.8333 for the test set.

3.3 | Ablation study

An ablation study is conducted to understand the effects of the different modules in our proposed framework. We

TABLE 1 Tuning of the hyperparameters of the proposed network (cyto-Knet).

Hyperparameter	Range/Values	Chosen value
l_r	$[10^{-i} \text{ for } i \in [2:6]]$	10^{-3}
Batch size	$[4, 8, 16, 32]$	16
$loss_{decode}$	[Dice, Cross-entropy]	Dice
$loss_{aux}$	[Dice, Cross-entropy]	Dice
$weight_{decode}$	$[i \cdot 10^{-1} \text{ for } i \in [2:9]]$	0.8
$weight_{aux}$	$[i \cdot 10^{-1} \text{ for } i \in [1:4]]$	0.4

Abbreviations: l_r, learning rate; $loss_{decode}$, loss function of the decode head; $loss_{aux}$, loss function of the auxiliary head; $weight_{decode}$, weight associated with $loss_{decode}$; $weight_{aux}$, weight associated with $loss_{aux}$.

TABLE 2 Quantitative performances of instance segmentation for construction (448 images) and test (50 images) sets.

Nuclei	Pixel-based metrics			Object-based metrics		
Subset	**Precision**	**Recall**	**F1 score**	**Dice**	**HD (pixels)**	**AJI**
Construction	0.8606	0.9089	0.8708	0.8916	8.56	0.7902
Test	0.8944	0.8828	0.8810	0.9048	10.20	0.8027
Cytoplasm	**Pixel-based metrics**					
Subset	**Precision**	**Recall**	**F1 score**			
Construction	0.8409	0.8941	0.8544			
Test	0.8747	0.8920	0.8779			
Plasma cells	**Pixel-based metrics**			**Object-based metrics**		
Subset	**Precision**	**Recall**	**F1 score**	**Dice**	**HD (pixels)**	**AJI**
Construction	0.8710	0.9219	0.8839	0.8953	27.18	0.8018
Test	0.9033	0.9049	0.8983	0.9130	28.84	0.8237

TABLE 3 Average computational times and performances on the test set for the different modules that compose our segmentation framework.

		Pixel-based metrics			Object-based metrics		
Method	**Time (s)**	**Precision**	**Recall**	**F1 score**	**Dice**	**HD (pixels)**	**AJI**
cyto-Knet	2.65	0.8108	0.9267	0.8571	0.8285	43.6065	0.7333
CS + cyto-Knet	2.88	0.8360	**0.9289**	0.8723	0.8631	36.4547	0.7633
CS + cyto-Knet + post	4.48	**0.9033**	0.9049	**0.8983**	**0.9130**	**28.8380**	**0.8237**

Note: The best performances are highlighted in bold.

Abbreviations: CS, color standardization (preprocessing); cyto-Knet, proposed segmentation network; post, postprocessing and segmentation refinement.

evaluate the results obtained using only the cyto-Knet, with and without the pre- and postprocessing stages. Table 3 shows the plasma cell segmentation results on the test set. Using only the deep learning model, we achieve an AJI of 0.7333. Adding the color standardization preprocessing step slightly improves the AJI to 0.7633. However, incorporating the full postprocessing strategy leads to a significant increase in AJI to 0.8237, a 6.04% improvement compared with the network alone.

3.4 | Comparison with state-of-the-art methods

In terms of a performance comparison with state-of-the-art methods, we evaluate the proposed segmentation strategy against several benchmarks:

1. The traditional K-Net,[21] which is integrated in our cyto-Knet model.
2. The ResNeSt model,[29] serving as the convolutional backbone for the top-performing method in the SegPC-2021 challenge.
3. The attention-based Deeplabv3+,[30] employed in the second-place solution of the challenge.
4. The Vision Transformer (ViT),[31] integrated into the method that claimed third place in the challenge.

We present a quantitative metrics comparison in Table 4, and a visual analysis is available in Figure 6 for a more comprehensive understanding of the results.

From Figure 6, it can be observed that the proposed cyto-Knet consistently estimates the correct number of plasma cells, while the state-of-the-art approaches incorrectly associate merged instances single cells and generate false-positive objects. The mean absolute percentage error (MAPE) on the number of detected plasma cells is computed for the test set, showing that our method achieves a MAPE of 12%, while the other approaches reach values of 41%, 31%, and 42% for ResNeSt, Deeplabv3+, and Vision Transformer, respectively.

Method	Pixel-based metrics			Object-based metrics		
	Precision	Recall	F1 score	Dice	HD (pixels)	AJI
K-Net	0.8139	0.9034	0.8478	0.7999	58.0502	0.6809
ResNeSt	0.8168	0.9157	0.8536	0.7953	50.5016	0.7134
Deeplabv3+	0.8526	0.8881	0.8598	0.8493	44.7354	0.7514
ViT	0.8019	**0.9227**	0.8491	0.8033	48.7715	0.7141
cyto-Knet	**0.9033**	0.9049	**0.8983**	**0.9130**	**28.8380**	**0.8237**

TABLE 4 Performance metric evaluation between the proposed approach (cyto-Knet) and the state-of-the-art methods on the test set. The best performances are highlighted in bold.

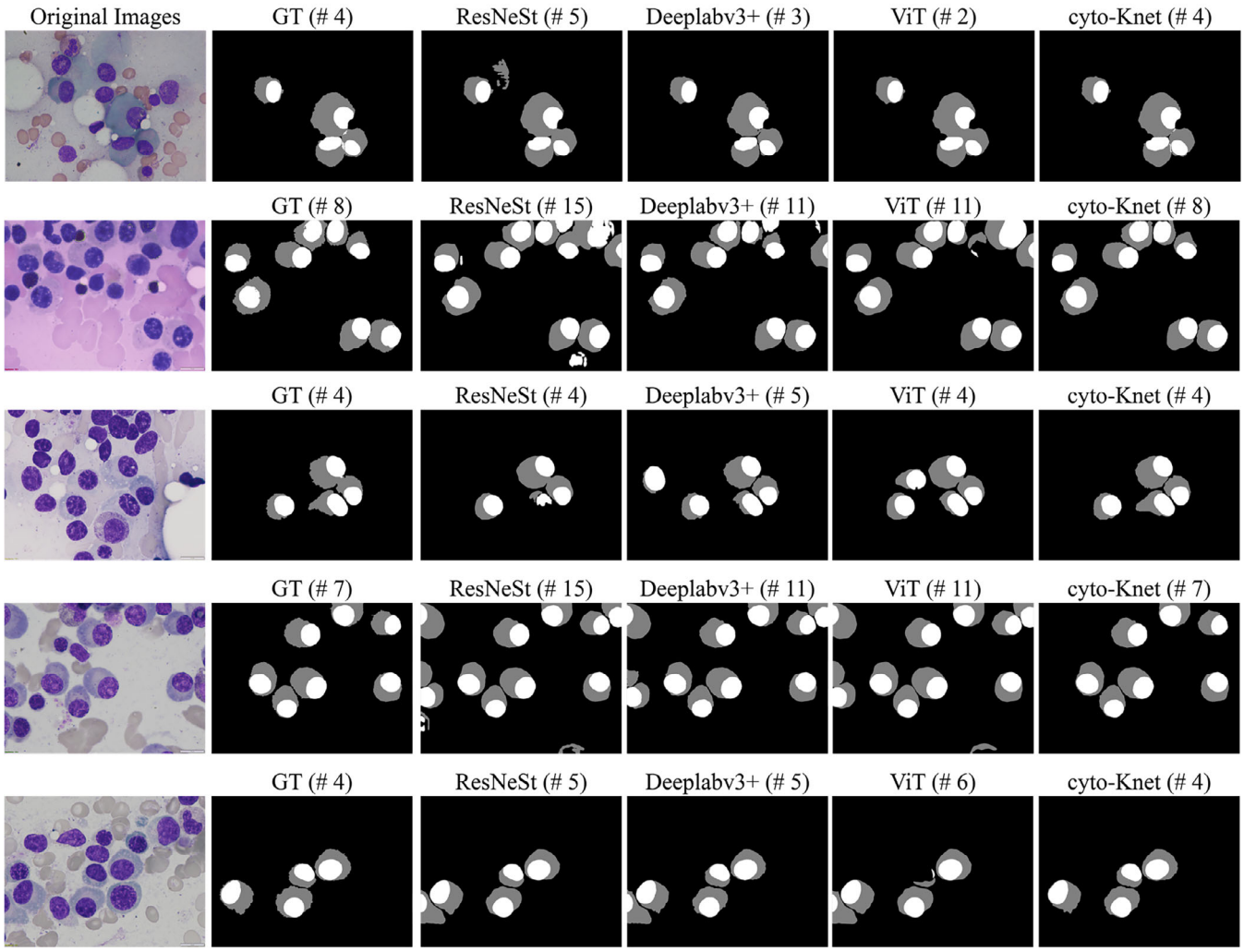

FIGURE 6 Sample images for state-of-the-art visual comparison. Original images and ground truth (GT) annotations are reported in the first and second columns, respectively; results of state-of-the-art approaches are shown from the third to the fifth column, and the proposed cyto-Knet segmentations are reported in the last column. The number of detected plasma cells is added in parenthesis for each method.

In addition, a recent study proposed a deeper and more compact split-attention U-Net, named DCSAU-Net,[11] and evaluated the model on a random test set of 49 images from the SegPC-2021 dataset. The authors reported the following pixel-based performance metrics: an accuracy of 0.950, a precision of 0.871, a recall of 0.910, and a F1 score of 0.886. The mIoU score was equal to 0.806. The proposed cyto-Knet outperforms DCSAU-Net for all metrics on the test set (accuracy: 0.967, precision: 0.903, recall: 0.905, F1 score: 0.898, mIoU: 0.833) except for the recall values that are comparable.

4 | DISCUSSION

The current gold standard for MM diagnosis involves manual analysis of suspicious plasma cells, also known as myeloma cells, from a serum sample taken through biopsy. However, this method can be time-consuming, prone to human errors, and affected by interoperator variability. The emergence of digital pathology has introduced new possibilities for pathologists. They can now inspect, label, and track digital tissue slides in high-resolution formats. This advancement enables easy sharing of second opinions and facilitates the integration of a laboratory information system to improve clinical workflow.[32] Additionally, digital cytology enables the use of computer-aided diagnosis (CAD) and artificial intelligence-based tools, enhancing the accuracy and the efficiency of MM diagnosis.

Cytological preparations present several challenges. These include irregular distribution of isolated and clustered plasma cells with different nuclei and cytoplasm sizes/shapes, as well as suboptimal staining processes that can negatively impact the chromatic appearance of cellular structures. To address these issues, we have employed the technique of instance segmentation, which focuses on pixel-wise prediction and object detection simultaneously. As illustrated in Figure 1, the instance segmentation task presents the following complexities: (i) suboptimal acquisition of the images, (ii) partially overlapping cells, and (iii) poorly defined contours of the objects of interest.

In this study, we propose a novel multiscale architecture (cyto-Knet), for instance segmentation of plasma cells from microscopic images of bone marrow aspiration. We have developed this framework using the publicly available SegPC-2021 dataset of myeloma cell images.[13] The proposed framework, summarized in Figure 2, combines deep learning techniques with heuristic approaches in the pre- and postprocessing steps. In the preprocessing stage, we introduce a novel color standardization method to balance different illuminant conditions. This method effectively calibrates the input cytological images and minimizes chromatic variability (Figure 3). Our approach successfully standardizes images captured from two different setups, such as microscope and cameras, and can be easily extended to other applications where white balancing of acquired images is not standard. The proposed cyto-Knet architecture (Figure 4) generates a softmax output indicating the probability of each pixel belonging to one of the four classes: background, nucleus, cytoplasm, and cell border. The introduction of a cell border class enables the subsequent postprocessing strategy to facilitate cell splitting, as depicted in the zoomed-in view of Figure 2. We employ the watershed algorithm with a rule-based heuristic suppression step to correctly separate merged objects, identifying only cells of interest and minimizing the risk of false-positive objects. Figure 5 graphically demonstrates the utility of the postprocessing strategy in removing single instances that are not cells of interest. The watershed transform used for separating merged instances detects 11 objects in the sample image. In the final mask, four plasma cells previously detected by the watershed algorithm are correctly discarded: cells #1, #3, and #10 have nuclei partially out of the field of view with an eccentricity above the threshold, and cell #6 consists of cytoplasm only (i.e., with a cytoplasm area ratio >85%). The postprocessing strategy relies on the watershed transform and a threshold-based heuristic approach. The employed watershed algorithm can be easily extended to other nucleated cells in the bone marrow aspirate images without requiring specific parameter tuning, thanks to the introduction of the new border class that facilitates the splitting of merged instances. The same principle applies to all thresholds used for removing unwanted objects, which are relative values based on softmax probability and the circular shape of cellular structures, such as nuclei and steatosis in other histological image data.

We conducted an extended performance validation by computing both pixel- and object-based quantitative metrics. These metrics allowed us to evaluate the accuracy of pixel-level misclassification, object-level detection errors, and both in a unified manner (e.g., AJI metric). Notably, this is the first study to report quantitative metrics, both pixel- and object-based, for the segmentation of all classes—nuclei, cytoplasm, and whole cells—as shown in Table 2. Our color standardization and postprocessing strategy proved effective in improving network performance, as demonstrated in the ablation study (Table 3). Specifically, the introduction of these steps resulted in a 4% increase in pixel-based F1 score and 9% increase in object-based AJI. Additionally, there was an improvement of approximately 15 pixels when evaluating contour matching between ground truth and automatic segmentations. The full pipeline with postprocessing requires 4.48 s, approximately 2 s longer than the network alone. However, this additional time is justified by the significant performance gain. The postprocessing also improves all pixel-based metrics. Precision, recall, and F1 score all see increases of 3%–4% with the full framework compared with the individual network alone.

Table 4 and Figure 6 demonstrate the performance comparison of our cyto-Knet framework with other state-of-the-art methods on the SegPC-2021 dataset. Our framework outperforms existing methods in terms of pixel-based F1 score, object-based AJI, and other evaluation metrics. These results highlight the effectiveness of

our proposed approach in accurately segmenting plasma cells in microscopic images of bone marrow aspiration. In the context of plasma cell counting, pathologists consider the percentage of abnormal plasma cells a crucial factor in the diagnosis of multiple myeloma. Our framework achieves the lowest mean absolute percentage error in counting these abnormal cells. Additionally, we have demonstrated that our method outperforms a recent study[11] that analyzed various state-of-the-art models and evaluated them on the same cytological dataset.

In summary, this work offers several benefits:

1. We present a hybrid framework that combines deep learning and heuristic techniques to achieve highly accurate instance segmentation of individual plasma cells in Giemsa-stained cytology images.
2. Our cyto-Knet includes an innovative color-balancing method as a preprocessing step, effectively standardizing the appearance of cytological images. This addresses issues related to variations in image quality resulting from different acquisition devices, such as microscopes or cameras.
3. Our network introduces a 4-class segmentation strategy with conditional kernels, significantly enhancing segmentation performance and accuracy.
4. We tackle the challenge of merged objects and the issue of false-positive object generation, employing a marker-based watershed algorithm in the postprocessing step.
5. Through extensive validation at both pixel- and object-based levels, we demonstrate superior performance compared with state-of-the-art techniques.

The main limitation of our study is directly related to the dataset employed. Our method is developed on the SegPC-2021 dataset,[13] which only includes annotations for cells identified as plasma cells by a senior onco-pathologist. As a result, not all visible cells were annotated, since not all visible cells are plasma cells of interest. Figure 7 shows sample images with low segmentation performance scores. Some sample images in our dataset exhibited low segmentation performance scores due to the presence of false-negative and false-positive objects. The false-negative objects were generated because some cells consisting only of the nucleus were annotated in specific images, unlike the majority of the dataset. The false-positive objects were a result of subtle differences in the manual annotations between stained plasma cells of interest and unstained cells considered part of the background. However, the challenge

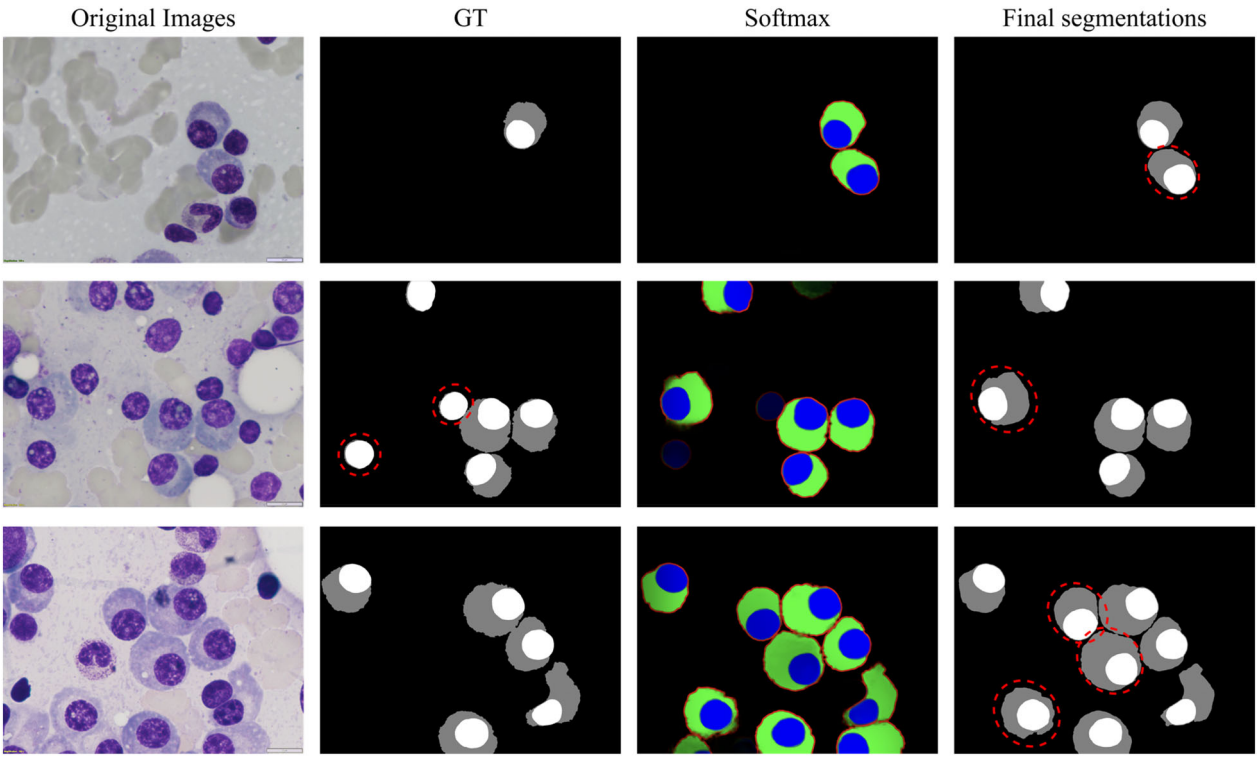

FIGURE 7 Sample images with faulty cases. Original images and ground truth (GT) annotations are reported in the first and second columns, respectively, while cyto-Knet softmax probability maps and final segmentations are reported in the third and fourth columns, respectively. False-positive and -negative objects are dashed in red.

organizers plan to release full annotations for a new dataset in the future, which will enable the construction of multiclass segmentation models. Additionally, our model was trained on a smaller number of images since the challenge's test set was not provided, which may have limited its performance.

5 | CONCLUSIONS

The current gold standard for diagnosis of multiple myeloma involves manual microscopic analysis of a serum sample taken by biopsy, which is time-consuming and affected by intra/interoperator variability. In this study, we have proposed a novel multiscale architecture, called cyto-Knet, for instance segmentation of plasma cells in Giemsa-stained cytological images. Our framework introduces several novel aspects, including a color-balancing algorithm as a preprocessing step, the integration of UPerNet into the K-Net architecture, the inclusion of a fourth class for cell boundary segmentation, and a post-processing algorithm to reduce false-positive segmentations. We further conduct an extended performance validation through the computation of both pixel- and object-based quantitative metrics using the publicly available SegPC-2021 dataset of myeloma cell images for the first time. The proposed cyto-Knet outperforms state-of-the-art approaches, by estimating the correct number of plasma cells, splitting merged instances correctly, and minimizing false-positive instance generation. In future studies, our proposed method could be integrated into an end-to-end system to support diagnostics. Additionally, the framework can be easily extended to other cytological stains and cellular modalities, such as multichromatic Papanicolaou staining, immunocytochemistry, or fluorescence microscopy.

AUTHOR CONTRIBUTIONS

Massimo Salvi and Filippo Molinari: involved in conceptualization. **Massimo Salvi, Nicola Michielli, Cristina Cattelino, Bruna Cotrufo, Matteo Giacosa, and Chiara Giovanzana**: involved in methodology. **Massimo Salvi, Nicola Michielli, and Kristen M. Meiburger**: involved in formal analysis and investigation. **Massimo Salvi and Nicola Michielli**: involved in data curation and validation. **Massimo Salvi and Nicola Michielli**: involved in writing—original draft preparation. **Cristina Cattelino, Bruna Cotrufo, Matteo Giacosa and Chiara Giovanzana, Kristen M. Meiburger, and Filippo Molinari**: involved in writing—review and editing. **Massimo Salvi, Kristen M. Meiburger, and Filippo Molinari**: involved in supervision. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors acknowledge support from Politecnico di Torino (Turin, Italy) through the Open Access initiative.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Massimo Salvi https://orcid.org/0000-0001-7225-7401
Nicola Michielli https://orcid.org/0000-0002-0301-637X

REFERENCES

1. Barlogie B, Shaughnessy J, Tricot G, et al. Treatment of multiple myeloma. *Blood*. 2004;103(1):20-32. doi:10.1182/blood-2003-04-1045
2. Rajkumar SV. Multiple myeloma: 2022 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2022;97(8):1086-1107. doi:10.1002/ajh.26590
3. Dogan EE, Arslan A, Demirel N, et al. Survival in patients with multiple myeloma: evaluation of possible associations with bone marrow fibrosis and investigation of factors independently associated with survival. *Bull Natl Res Centre*. 2022;46(1):242. doi:10.1186/s42269-022-00926-6
4. Saeedizadeh Z, Mehri Dehnavi A, Talebi A, Rabbani H, Sarrafzadeh O, Vard A. Automatic recognition of myeloma cells in microscopic images using bottleneck algorithm, modified watershed and SVM classifier. *J Microsc*. 2016;261(1):46-56. doi:10.1111/jmi.12314
5. Vyshnav MT, Sowmya V, Gopalakrishnan EA, Variyar VVS, Menon VK, & Soman KP. Deep learning based approach for multiple myeloma detection. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2020, 1–7. 10.1109/ICCCNT49239.2020.9225651
6. Hafiz AM, Bhat GM. A survey on instance segmentation: state of the art. *Int J Multimedia Inform Retriev*. 2020;9(3):171-189. doi:10.1007/s13735-020-00195-x
7. Bozorgpour A, Azad R, Showkatian E, Sulaiman A. Multi-scale regional attention Deeplab3+: multiple myeloma plasma cells segmentation in microscopic images. 2021. doi:10.48550/arXiv.2105.06238
8. Azad R, Bozorgpour IA, Asadi-Aghbolaghi M, Merhof D, & Escalera S. Deep frequency Re-calibration U-net for medical image segmentation. 2021 IEEE/CVF International Conference

on Computer Vision Workshops (ICCVW). 2021, 3267–3276. 10.1109/ICCVW54120.2021.00366
9. Pandey D, Gupta P, Bhattacharya S, Sinha A, Agarwal R. Transformer assisted convolutional network for cell instance segmentation. 2021. doi:10.48550/arXiv.2110.02270
10. Faura ÁG, Štepec D, Martinčič T, Skočaj D. Segmentation of multiple myeloma plasma cells in microscopy images with noisy labels. Proc SPIE, 12033, 120330N 2022. doi:10.1117/12.2607458
11. Xu Q, Ma Z, He N, Duan W. DCSAU-net: a deeper and more compact split-attention U-net for medical image segmentation. *Comput Biol Med*. 2023;154:106626. doi:10.1016/j.compbiomed.2023.106626
12. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans Med Imaging*. 2017;36(7):1550-1560. doi:10.1109/TMI.2017.2677499
13. Gupta A, Gehlot S, Goswami S, et al. SegPC-2021: a challenge & dataset on segmentation of multiple myeloma plasma cells from microscopic images. *Med Image Anal*. 2023;83:102677. doi:10.1016/j.media.2022.102677
14. Gehlot S, Gupta A, & Gupta R. EDNFC-net: convolutional neural network with nested feature concatenation for nuclei-instance segmentation. ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020, 1389–1393. 10.1109/ICASSP40776.2020.9053633
15. Gupta A, Duggal R, Gehlot S, et al. GCTI-SN: geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images. *Med Image Anal*. 2020;65:101788. doi:10.1016/j.media.2020.101788
16. Gupta A, Mallick P, Sharma O, Gupta R, Duggal R. PCSeg: color model driven probabilistic multiphase level set based tool for plasma cell segmentation in multiple myeloma. *PloS One*. 2018;13(12):e0207908. doi:10.1371/journal.pone.0207908
17. Pinco J, Goulart RA, Otis CN, Garb J, Pantanowitz L. Impact of digital image manipulation in cytology. *Arch Pathol Lab Med*. 2009;133(1):57-61. doi:10.5858/133.1.57
18. Lam EY. Combining gray world and retinex theory for automatic white balance in digital photography. Proceedings of the Ninth International Symposium on Consumer Electronics, 2005. (ISCE 2005). 2005, 134–139. 10.1109/ISCE.2005.1502356
19. Provenzi E, Gatta C, Fierro M, Rizzi A. A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast. *IEEE Trans Pattern Anal Mach Intell*. 2008;30(10):1757-1770. doi:10.1109/TPAMI.2007.70827
20. Xiao T, Liu Y, Zhou B, Jiang Y, & Sun J. Unified perceptual parsing for scene understanding. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11209 LNCS. 2018, 432–448. 10.1007/978-3-030-01228-1_26
21. Zhang W, Pang J, Chen K, Loy CC. K-net: towards unified image segmentation. *Advances in Neural Information Processing Systems*. 2021;34:10326-10338.
22. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, & Guo B. Swin transformer: hierarchical vision transformer using shifted windows. Proceedings of the IEEE International Conference on Computer Vision. 2021, 9992–10002. 10.1109/ICCV48922.2021.00986
23. Salvi M, Mogetta A, Gambella A, et al. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. *Comput Med Imaging Graph*. 2021;90:101930. doi:10.1016/j.compmedimag.2021.101930
24. Salvi M, Mogetta A, Meiburger KM, et al. Karpinski score under digital investigation: a fully automated segmentation algorithm to identify vascular and stromal injury of Donors' kidneys. *Electronics*. 2020;9(10):1644. doi:10.3390/electronics9101644
25. Huang H-Q, Fang X-Z, Shi J, Hu J. Abnormal localization of immature precursors (ALIP) detection for early prediction of acute myelocytic leukemia (AML) relapse. *Med Biol Eng Comput*. 2014;52(2):121-129. doi:10.1007/s11517-013-1122-x
26. Kornilov AS, Safonov IV. An overview of watershed algorithm implementations in open source libraries. *J Imaging*. 2018;4(10):123. doi:10.3390/jimaging4100123
27. Salvi M, Molinari F. Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images. *Biomed Eng Online*. 2018;17(1):89. doi:10.1186/s12938-018-0518-0
28. Huttenlocher DP, Klanderman GA, Rucklidge W. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 1993;15:850-863.
29. Zhang H, Wu C, Zhang Z, Zhu Y, Lin H, Zhang Z, Sun Y, He T, Mueller J, Manmatha R, Li M, & Smola A. ResNeSt: Split-attention networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2022, 2735–2745. 10.1109/CVPRW56347.2022.00309
30. Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S. Attention Deeplabv3+: multi-level context attention mechanism for skin lesion segmentation. In: Bartoli A, Fusiello A, eds. *BT – Computer Vision – ECCV 2020 Workshops*. Springer International Publishing; 2020:251-266.
31. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. 2020. doi:10.48550/arXiv.2010.11929
32. Dey P. Digital pathology. *Basic and Advanced Laboratory Techniques in Histopathology and Cytology*. Springer Nature Singapore; 2022:195-203. doi:10.1007/978-981-19-6616-3_18

How to cite this article: Salvi M, Michielli N, Meiburger KM, et al. cyto-Knet: An instance segmentation approach for multiple myeloma plasma cells using conditional kernels. *Int J Imaging Syst Technol*. 2024;34(1):e22984. doi:10.1002/ima.22984