WAS Science Nature
WASSN

ISSN: 2766 - 7715

WAS PRESS

Article

A comparison of U-net backbone architectures for the automatic white blood cells segmentation

Mohammed Hakim BENDIABDALLAH [1,2]* and Nesma SETTOUTI [1]

[1] Biomedical Engineering Laboratory, Faculty of Technology, University of Tlemcen, Algeria.
[2] University of Ain Temouchent Belhadj Bouchaib, Ain-Temouchent, Algeria.
* Correspondence: hakim.bendiabdellah@univ-tlemcen.dz

Received: 08-03-2021; Accepted: 04-05-2021; Published: 04-09-2021

Abstract: Reliable recognition of white blood cells is an essential step in the diagnosis of several types of cancer. Therefore, the segmentation of white blood cells plays an essential role and is an important part of the medical diagnostic system. Manual cell diagnosis involves doctors visually examining microscopic images to detect any cellular abnormalities. This step is costly and time-consuming. An automated system based on white blood cell identification provides a more accurate result than the manual method. Image segmentation is one of the crucial contributions of a deep learning community to the medical field. In this paper, we demonstrate how the U-Net type architecture can be improved by the use of the pre-trained encoder, a comparison of several efficient methods for automatic recognition of white blood cells using the original U-NET, different pre-trained classification networks are used as the backbone to obtain better performance. The architecture of RESNET-50 obtains the best segmentation results on testing data for automatic recognition in cytological images with a less amount of training epochs.

Keywords: white blood cells segmentation, deep Learning, transfer learning, U-NET, Loss Function, cytological image's dataset.

1. Introduction

The expertise of spinal cord smears represents the cornerstone of hematological diagnosis. Indeed, the bone marrow is made up of stem cells from which blood cells (WBC) are produced, and in the event of an abnormality in one of the components of the blood (in the event of a deficiency or proliferation) the cells of the bone marrow can in be the cause. Unlike the blood smear, it is sufficient to focus on a microscopic field which seems adequate (many cells, well spread out and well stained) and to carry out the count of all the cells present in this field; then move to another field that seems adequate, and so on **[1]**. Clearly, this test is an important indicator in the detection of certain blood abnormalities. Blood morphology is made up of three elements: cells such as red blood cells (erythrocytes) and white blood cells (leukocytes) as well as blood platelets (not considered cells). The expression of the shape and number of white blood cells (WBC) has many quantitative and

informative clues **[2]**. For example, increasing or decreasing white blood cells is very critical and may receive medical attention.

Within the framework of medical image analysis techniques, the segmentation of white blood cells is a key problem that has been highlighted in this work. The segmentation of microscopic images uses information from the image (colour, grayscale and spatial) to delineate different anatomical structures, including white blood cells (WBC) which are made up of nucleus and cytoplasm.

In this work, a deep learning approach has been suggested using U-Net to solve the problem of automatic recognition in cytological images, eight pre-trained encoder networks of U-Net models were appraised in the same experimental environment and with the same data. The distinction between white blood cells (leukocytes) in microscopic images of bone marrow and peripheral blood allows for accurate diagnoses of different cancers using a cost-effective method for fast, reliable, and efficient detection of nucleus and cytoplasm, which are clinically very important. Firstly, the most powerful structure for the encoder of U-Net is exposed in comparison to multiple deep learning models. Secondly, semantic segmentation with the best model is performed, then we show the performance of the best model with examples. Finally, a comparison with other models is made.

The paper is organized as follows: semantic segmentation related works of blood stem cells from bone marrow are presented in section 2. By next, materials and principal methods applied are defined in section 3. In section 4, we describe the cytological images dataset used in this study, with the experimental setting for the comparative study. In section 5 experimental results are discussed. Finally, conclusions from this study and possible future works are presented in section 6.

2. Related works

Several research works have been carried out on the semantic segmentation of blood stem cells from bone marrow **[3–5]**. In this work, previous works and methods are cited that have been proposed and applied to the real image of the cytological image's dataset **[6]** on which our study is based. These microscopic images of blood stem cells of the bone marrow were collected within the Haemobiology Service, University Hospital Center of Tlemcen, Algeria, on slides staining type MGG (May Grunwald Giemsa). Research works in the literature can be divided into four segmentation approaches:

Morphological approaches:

Benazzouz et al. [**6**] proposed an automated identification of plasma cells in bone marrow images. The steps of their segmentation model were divided into two phases. The first one used Otsu thresholding to extract the nucleus (green label) and the second one used the region growing on the obtained nucleus to delineate the cytoplasm. After segmentation, a classification of the obtained globules is used to count them. This method showed promising results when extracting the white blood cell named Leucocyte. Besides, segmentation of bone marrow images based on the Watershed transformation was proposed by *Baghli et al.* **[7]**. Its principle is to consider the image as a topographic map where the user must define starting points for the algorithm to flood the basins (objects to be detected) until there is a meeting point between the different basins (regions). Then the regions are merged by integrating the uncertainty on the colour through the theory of evidence. In the end, the

classification of the obtained globules is performed by three classifiers: SVM, Knn, and the decision tree. A Multi Features Based Approach for White Blood Cells Segmentation and Classification in Peripheral Blood and Bone Marrow Images was proposed by *Benomar et al*. **[8]**. It is a system that allows segmentation and differential counting of white blood cells, the process begins by highlighting WBCs by the stretch decorrelation method. Then, Otsu's thresholding is applied to the edited image using a colour transformation. Segmentation is performed by Watershed to determine the boundaries of the blood cells followed by cleaning the image to remove false positives. At the end of this segmentation, a colour scheme is used to separate the nucleus from the cytoplasm.

Pixel-based Classification approaches:

Pixel-based classification involves classifying each pixel in the image to a region class by machine learning approaches. Indeed, these methods perform a separation in the characteristic space of each pixel, the representation space can be constituted by the color or texture information so that a projection in this space achieves linear or non-linear boundaries between regions of the image. For white blood cell segmentation, *Settouti et al.* **[9]** proposed an automatic method based on region growth by classifying neighbouring pixels from the pixels of interest in the image with minimal intervention by the expert. The points of interest are detected by the ultimate erosion morphological operator and two classifiers are applied for classification: Decision Tree (mono-classifier) and Random Forest (Multi-classifiers / ensemble method). The main limitation of this method is the long processing time, which makes it useless in a big data problem. In this case, two solutions have been proposed to resolve this problem, which is: Involving instance selection algorithms for a pixel reduction process that can reduce the cost of storing and computing image segmentation by selecting relevant pixels to the pixel-based classification task. *Saidi et al.* **[10]** proposed the EMIS Algorithm, an instance selection approach based on ensemble methods that use the ensemble margin as a selection criterion to overcome the problem of sensitivity to noise. Subsequently, in addition to this time saving, in another work, *Settouti et al.* **[11]** identify the relevant colour spaces, which provide more information in the WBC segmentation process and eliminate the redundant and unnecessary characteristics of all images feature extraction. They proposed the IVsel algorithm "Instance & Variable selection", which highlights the importance of selecting only the most useful instances and variables to separate the different ROIs.

Super-pixel-based Classification approaches:

The current trend is towards the application of super-pixel classification which has great potential in the segmentation of colour images in the segmentation process. It is a clustering technique that allows the image to be subdivided into *k* homogeneous clusters allowing to accelerate and improve the quality of segmentation. *Bechar et al.* **[12]** developed a segmentation procedure based on super-pixel classification, where characterization based on image colour information is done at the super-pixel level. They performed different ways of characterization to study the influence of colour normalization, colour information, and characterization technique on the segmentation results of white blood cells. The findings indicate that colour normalization provides characterization precision and significant segmentation improvements. Besides, *Bechar et al*. **[13]** demonstrated the application potential of semi-supervision in the segmentation of cytological images and the recognition of white blood cells. A comparison is carried out between the multi-classifiers and the mono-

classifier in supervised mode (random forests vs. Decision tree) and semi-supervised mode (co-FOREST vs. Self-learning SETRED), the application of algorithms semi-supervised learning have shown their superiority over supervised learning.

Deep Learning approaches:

The advent of deep learning has eased the heavy lifting of the physician by making it possible to make accurate diagnoses in a short time. Particularly, Convolution Neuronal Networks (CNNs) have proven to be very robust in solving image classification problems and several architectures have been proposed to increase their performance, notably for microscopic images of the blood cells of the bone marrow. The success of deep learning methods in performing image classification has extended their use to solve more complex tasks including semantic segmentation by interpreting it as either a regression or discrimination problem. For the same case application of our work, recently, *Khouani et al.* **[14]** have conducted a study of deep learning methods for the automatic segmentation of regions of the nucleus and cytoplasm in cytological images. The proposed model is based on the use of Mask R-CNN with an improvement in the architecture and the stages of pre- and post-processing. The results obtained are very promising and show the power of deep learning methods in the field of image processing.

In summary, as with the previously mentioned approaches, classical segmentation methods have major limitations that do not favour their deployment in clinics to perform critical tasks. If an approach is efficient, it requires a significant amount of execution time or human interaction. If the approach is automatic and does not require a learning mechanism, it is very sensitive to noise in the image. For methods requiring a learning mechanism, a detection phase is almost mandatory because the performance of the segmentation is closely linked to it. Moreover, the choice of the features to be extracted is problematic because even this restricts the scope of the method to a specific type of image where the classifier learns a feature of the structure to be segmented and not its global representation. This also limits the application of this type of method in the case where the anatomical structures are deformed. The success of deep learning (DL) lies mainly in its deference to traditional machine learning (ML) approaches. Indeed, ML models improve gradually but lack precision so the user must guide him by solving the problem explicitly, unlike DL which does it and can dispense with the feature extraction phase. Deep Learning has undergone a revolution over the past few years and an infinite number of architectures and models have been produced. Most of these models gave more than satisfactory results.

Through the literature review, it was noted that the semantic segmentation of blood stem cells from bone marrow arouses a lot of interest within the scientific community. Classical approaches have provided initial solutions for WBC segmentation but with the shortcomings associated with each of the methods employed. Deep learning-oriented approaches have eliminated most of these limitations, but it nevertheless has several interrogations. First, the different architectures and models that exist make it difficult to choose the right network. Second, the hyperparameters of the network are difficult to evaluate a priori. Indeed, the number of layers, the number of neurons per layer, or the different connections between layers are crucial elements and essentially determined by a good intuition or by a succession of tests/calculation of errors (which is costly in time). The number of training samples is also a determining factor, and it often happens that this is too small compared to the number of parameters (weight) of the network. There are solutions

such as artificially increasing their number or even reducing the number of free parameters (by pre-learning the first layers for example).

In this paper, we present a comparison of several efficient methods for automatic recognition of white blood cells using the U-Net convolutional neural network, and different famous architecture: MobileNet, MobileNet V2, RESNET 18, RESNET 34, RESNET 50, RESNET 152, VGG 16, VGG 19 as the backbone to find the best network with the best trade-off between performances/quality of segmentation and training parameters of each architecture.

3. Materials and Methods

There are several deep learning architectures that can solve this semantic segmentation problem. The general semantic segmentation network consists of an encoder and a decoder, U-Net architectures are normally considered as one of the most powerful tools for segmentation of biomedical images. Different pre-trained CNN architectures are used as backbones of U-Net, and passing them to the U-Net decoder, the process is illustrated in Figure 1.

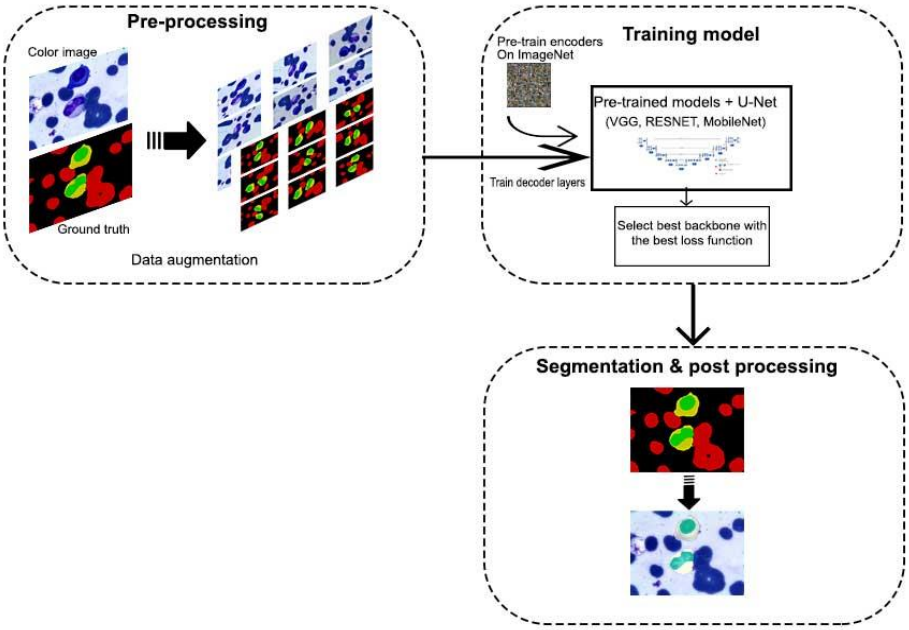


Figure 1. Diagram of the proposed approach for automatic segmentation using U-Net backbones.

3.1. The model:

U-net **[15]** is an optimized semantic segmentation network based on Fully Convolutional Network (FCN). The architecture consists of a contracting path (encoder) to capture context and a symmetric expanding path (decoder) that enables precise localization and improves performance on segmentation tasks. This model is chosen because the emergence of U-Net has brought great prospects for deep learning in the field of medical image analysis. It builds on previous convolutional networks to work more efficiently with fewer training images and to achieve more efficient segmentation. Batch normalization is used before convolutional layers for internal covariate shift to achieve better performance and accelerate convergence **[16].** Also, a dropout layer is utilized in the structure to reduce over-fitting problems.

3.2. Data augmentation

U-Net is capable of learning from a relatively small training set. In most cases, data sets for image segmentation consist of at most thousands of images, since manual preparation of the masks is a very costly procedure, so data augmentation is used in this case. Data augmentation is essential to teach the network invariance and robustness properties. Using our small dataset of images and masks, new images can be generated that will be as insightful and useful to our model as our original images. Random transformations on the input images were used in the database augmentation. Inputs are randomly rotated vertically and horizontally, then zoomed them. The brightness of the images is also randomly increased and decreased because the colour variation in the images was a significant segmentation complication due to the quality of the captor used during image capturing.

3.3 Transfer learning technique and training models:

Transfer learning is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision processing tasks. It reduces training time considerably and leads to effective models **[17,18]** even with a small training set like ours.

As illustrated in Figure 1, MobileNet **[19]**, VGG **[20],** RESNET **[21]**) deep neural network are used as backbones (an encoder) in our U-Net network, the residual blocks in RESNET with skip connections helped in making a deeper and deeper convolution neural network and achieved record-breaking results for classification on the ImageNet dataset. All mentioned backbones weights are pre-trained on The ImageNet data set **[22]**, the advantage of doing so is to shorten the learning procedure, which is done by the last few layers of the network, to speed up convergence and to achieve high performance as compared to a non-pre-trained model. These backbones are used as the first half (encoder) of U-net **[23]**. Then, the decoder layers are trained with the augmented dataset. It helps save the training procedure and enhances the advantage of U-net's ability to learn from small data.

During the convolutional neural network training, validation is used to detect when overfitting starts, training is stopped when performance on the validation set starts to degrade in order to avoid the overfitting on the training data ("early stopping") **[24]**.

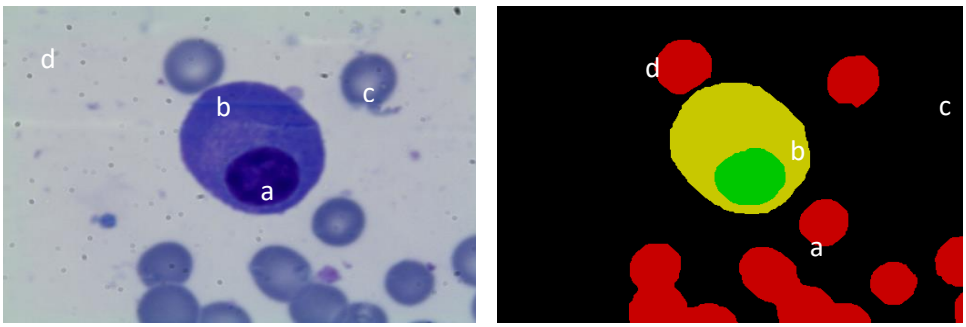
4. Results

4.1. The cytological images dataset

The dataset belongs to the field of hematology. It contains microscopic images of the blood cells of the bone marrow which were collected at the Haemobiology Service, University Hospital Center of Tlemcen, Algeria by *Benazzouz et al*. **[6].** This dataset is acquired in the LEICA environment (camera and microscope), the characteristics of the database are described in Table 1. Figure 2 shows some images of this dataset in the top and their Ground Truth (bottom) where the expert selects 4 regions: nucleus, cytoplasm, red blood cells and plasma.

Table 1. The cytological images dataset features **[6]**.

Number of Images	Original size	Format	Magnification
87	1024x768	BMP	X100



(**A**) (**B**)

Figure 2. Samples of cytological images dataset (A) with the ground truth (B), where: (a) Nucleus, (b) cytoplasm, (c) red cell and (d) plasma **[6]**.

The protocol for obtaining these images is known as the spinal cord smear (myelogram). It consists of spreading bone marrow on a microscopic slide to be able to study the morphology of the cells present as well as their number after being stained and fixed using the microscope. 80% of the images are randomly chosen for training and validation, then the rest for testing. For the optimal configuration for our machine we have resized the images to 256x256.

4.2. Experimental setting

All experiments are performed on a computer with GeForce NVIDIA GTX 1060 graphics cards. The proposed network models were implemented with python and TensorFlow v 2.2.0. The models are trained using **Adam gradient-based optimization algorithm** with step decay of learning rate, the initial learning rate is 1e-3, it is a popular algorithm in the field of deep learning because is computationally efficient and has small memory requirements **[25]**.

The early stopping criterion is determined at 10 training epochs both in the training and validation phase for each architecture.

Loss function:

Since the image segmentation task is considered as a pixel classification problem, the choice of the loss function is very important while designing complex image segmentation. Both cross-entropy (CE) eq. (1) and Dice loss (DL) eq. (2) functions are used, then compared their performances to find the best metric with our database.

Categorical Cross Entropy loss **[26]**:

$$CE(y, \hat{p}) = -(\mathrm{y}\log(\hat{p}) + (1 - \mathrm{y})\log(1 - \hat{p})) \quad (1)$$

Here, $\hat{p}$ is the predicted value by the prediction model and y is the ground truth.

Dice Loss **[27]**:

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p} + 1}{y + p + 1} \quad (2)$$

Here, 1 is added in numerator and denominator to ensure that the function is not undefined in edge case scenarios such as when $y = \hat{p} = 0$.

4.3. Evaluation Metrics

Several experiments are performed using data augmentation and Transfer learning for different encoders pre-trained on ImageNeT (MobileNet, MobileNet V2, RESNET 18, RESNET 34, RESNET 50, RESNET 152, VGG 16, VGG 19). Due to the performance limitations of the used machine, all images are resized to 256x256, this did not significantly affect the quality of segmentation.

The classification performances are evaluated based on the Precision eq. (3) and the most used metrics for semantic segmentation F1 Score (Dice Coefficient) eq. (4), sensitivity (5), specificity (6) and accuracy (7) metrics are also used in Table 4.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F - score = Dice\ Coefficient = 2 * \frac{Precision\ .\ Recall}{Precision + Recall} = 2 * \frac{TP}{2 * TP + FP + FN} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Where:

- True positives (TP): The intersection between segmentation and ground truth,
- False positives (FP): Segmented parts not overlapping the ground truth,
- False negatives (FN): Missed parts of the ground truth,
- True negatives (TN): Part of the image beyond the union between segmentation and ground truth.

Table 2. Performances, Epochs using Dice loss and Cross-entropy loss for each backbone.

Encoder	Metrics	Nucleus		Cytoplasm		Epochs	
		CrossEntr Loss	Dice Loss	CrossEntr Loss	Dice Loss	CrossEntr Loss	Dice Loss
U-NET	**Precision**	0.9798	0.9796	0.8853	0.8754	53	43
	F-score/Dice	0.9547	0.9496	0.9271	0.9192		
	Accuracy	0.9974	0.9971	0.9930	0.9922		
MOBILNET	**Precision**	0.9360	0.9863	0.8852	0.8639	14	23
	F-score/Dice	0.9434	0.9406	0.9152	0.9192		
	Accuracy	0.9967	0.9967	0.9920	0.9921		
MOBILENET V2	**Precision**	0.9936	0.9562	0.8600	0.8672	29	49
	F-score/Dice	0.9360	0.9450	0.9173	0.9123		
	Accuracy	0.9965	0.9968	0.9919	0.9916		
RESNET 50	**Precision**	0.9800	0.9900	0.8891	0.8816	**13**	37
	F-score/Dice	**0.9577**	0.9419	**0.9314**	0.9214		
	Accuracy	0.9976	0.9967	0.9934	0.9925		

VGG 19	**Precision**	0.9664	0.9806	0.8877	0.8703	38	52
	F-score/Dice	0.8154	0.9282	0.9206	0.9178		
	Accuracy	0.9906	0.9960	0.9925	0.9921		
VGG 16	**Precision**	0.9105	0.9701	0.8872	0.8554	27	69
	F-score/Dice	0.8707	0.9205	0.9010	0.9079		
	Accuracy	0.9927	0.9956	0.9908	0.9910		
RESNET 152	**Precision**	0.9440	0.9878	0.8934	0.8720	24	44
	F-score/Dice	0.9477	0.9400	0.9226	0.9231		
	Accuracy	0.9969	0.9966	0.9927	0.9925		
RESNET 34	**Precision**	0.9754	0.9878	0.8803	0.8814	16	41
	F-score/Dice	0.9485	0.9493	0.9252	0.9284		
	Accuracy	0.9971	0.9971	0.9928	0.9931		
RESNET 18	**Precision**	0.9202	0.9807	0.8921	0.8871	22	36
	F-score/Dice	0.9269	0.9435	0.9111	0.9284		
	Accuracy	0.9957	0.9968	0.9917	0.9932		

Bold numbers represent the best results using F-Score

5. Discussion

After analysing the results in Table 2, the first observation is that the accuracy is almost equal to 1 using almost all encoders, this metric can sometimes provide misleading results in segmentation when the representation of white blood cells is small in the image. , this problem is called class imbalance because the measurement will be biased mainly reporting how the negative case is identified (plasma and red blood cells). This is intended to illustrate the fact that high pixel accuracy does not always imply greater segmentation capacity. Therefore, two other metrics are used: Precision and F-Score (Dice score) which are better able to handle this problem.

Results in Table 2 demonstrated that the use of Standard U-Net and RESNET-50 gives the best performance, but RESNET-50 converges to better results in only 13 epochs using the Cross-entropy loss function, Dice score has achieved a promising result for nucleus and cytoplasm segmentation: **0.957** and **0.931** respectively. The early stopping criterion is determined at 10 training epochs for each architecture if the loss function does not improve after 10 iterations, because too many epochs can lead to overfitting of the training dataset.

Table 3. Number of training parameters for each architecture.

Encoder	**U-NET**	**MOBILNET**	**MOBILENET V2**	**RESNET 50**	**VGG 19**	**VGG 16**	**RESNET 152**	**RESNET 34**	**RESNET 18**
Total params	31,055,748	8,336,772	8,047,876	32,561,549	29,062,404	23,752,708	67,295,629	24,456,589	14,341,005

As shown in Table 3, the number of learning parameters plays an important role, RESNET-50 has a deeper architecture and a larger number of parameters, so it is obvious that it achieves better performance as compared to MobileNet, ResNet-18, and ResNet-34. However, RESNET-152 has a very deep architecture so more parameters should achieve better performance in theory, but the experimental results are just the opposite, because of the size of the database which is too small. Here we got a preliminary conclusion: The depth of the network is the most crucial point to cell segmentation. On the other hand, Table 2

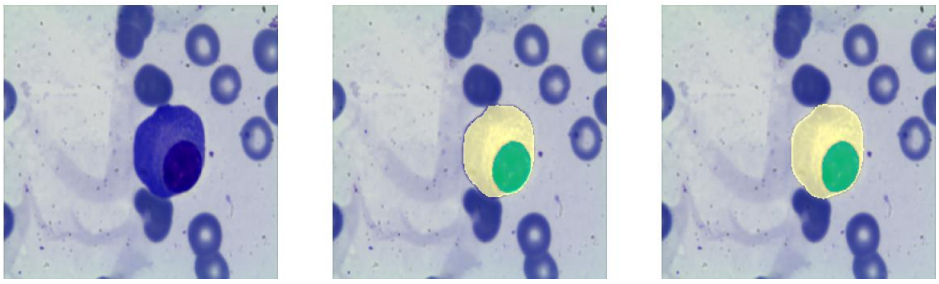
demonstrated that the model with Resnet encoder showed better results than VGG-16 and VGG-19, although they all have a large number of parameters. This may be due to the fact that ResNet in its architecture allows the model to learn more efficiently using skip connections and its greater depth compared to VGG allows it to capture more complex concepts which makes it a more efficient encoder.

Table 4. Performances, using Dice loss and Cross-entropy loss for RESNET-50 encoder.

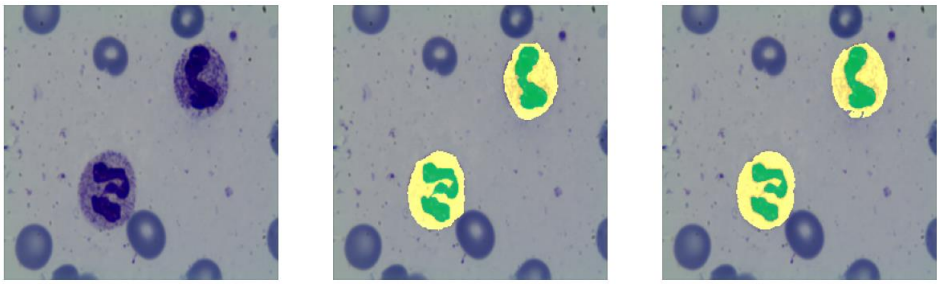
	Nucleus		**Cytoplasm**	
	CrossEntr Loss	**Dice Loss**	**CrossEntr Loss**	**Dice Loss**
Precision	**0.9800**	**0.9900**	**0.8891**	0.8816
Sensitivity	**0.9365**	**0.8982**	**0.9780**	0.9651
Specificity	**0.9994**	**0.9997**	**0.9942**	0.9938
F-Score/Dice	**0.9577**	**0.9419**	**0.9314**	0.9214

The specificity using RESNET-50 pre-trained encoder (Table 4) is practically equal to 1, this means the system avoid false alarms, the clinician is more assured in the automatic system. It is also noted that the use of the cross-entropy loss function on this database has allowed a clear improvement in the true positive rate, which allowed a better segmentation in the majority of the test images, as illustrated on some samples in Figure 3.

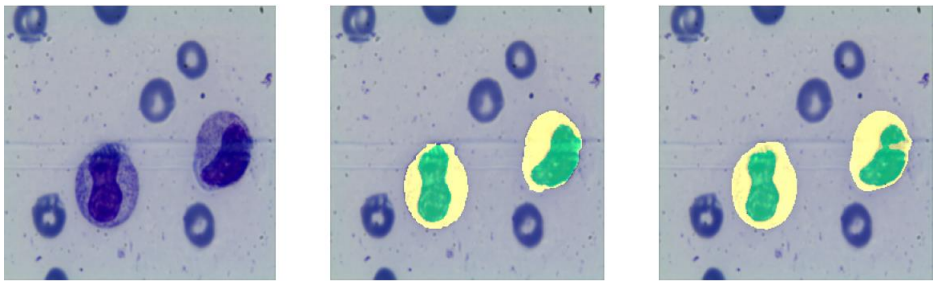
(Image 1)



(Image 2)



(Image 3)



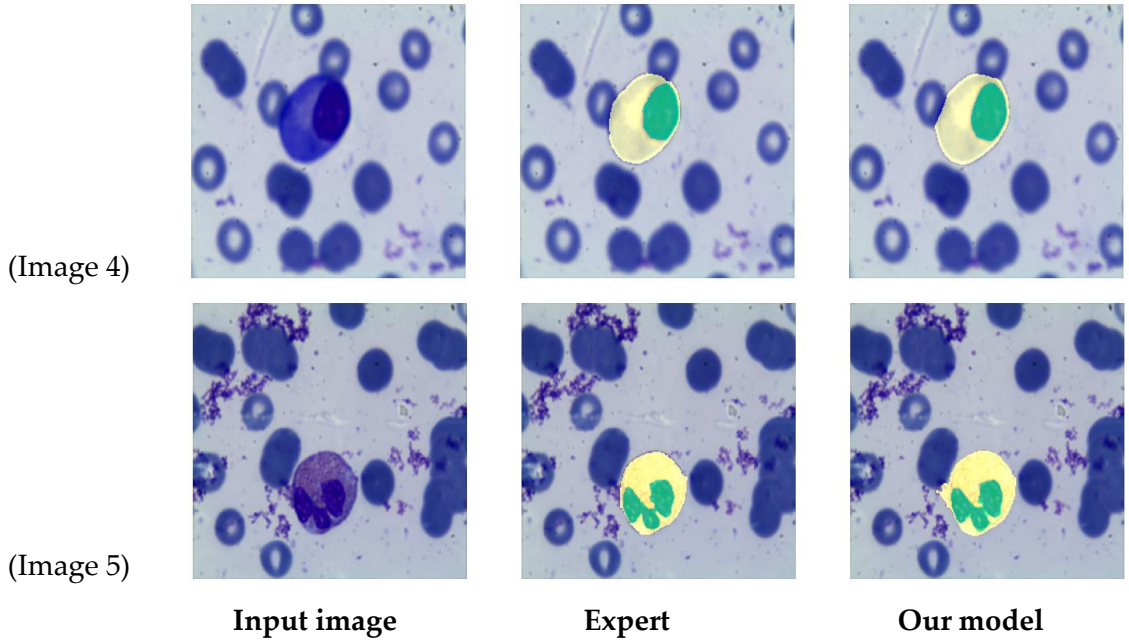


Figure 3. Results of automatic segmentation by U-NET using RESNET-50 Encoder

Five images are randomly selected from the test database (Figure 3) to discuss the performance of our semantic segmentation model. The mask is overlapped on the original image to have better visibility of the segmentation quality. We note that the separation of cytoplasm and nuclei is almost identical to the expert's labelling, also the model can distinguish white blood cells from red blood cells, for example in Image 4 (Figure 3).

We find that the colour variation in the images does not affect the segmentation, this is partly due to data augmentation and to be precise, the fact of increasing and decreasing the brightness. These results can affirm the effectiveness and great speed of response compared to other methods (small number of epochs) using transfer learning (with pre-trained weights on ImageNet). The use of RESNET-50 as backbone gives the best results, we can explain this by using a deeper convolution network through series of residual blocks (with skip connections) to help us addresses the vanishing gradient problem.

Having demonstrated the superiority of the RESNET-50 as a backbone model over other models, in Table3, we will now briefly compare the results we obtained in this study with those of previous works.

Table 5. Comparison with related works on WBC segmentation method with the same dataset.

Authors	**Model/Algorithm**	**Nucleus**		**Cytoplasm**	
		Precision	**F-Score/Dice**	**Precision**	**F-Score/Dice**
Benazzouz et al. **[6]**	Otsu + classification	95.02	-	84.53	-
Baghli et al. **[7]**	Evidence theory	95.90	-	88.4	-
Benomar et al. **[8]**	Otsu + watershed	96.87	-	92.50	-
Settouti et al. **[9]**	Pixel-based approach	**99.12**	0.9532	**97.15**	0.8982
Saidi et al. **[10]**	EMIS	99.05	0.88	95.05	0.61
Settouti et al. **[11]**	IVsel	99.10	0.84	94.99	0.4518
Our Approach	U-Net RESNET-50	98.00	**0.9577**	88.91	**0.9314**

If we compare the performances of the quoted works in Table 5 (the comparison is based only on precision and f-score, as the majority of previous works have provided only these metrics), we can establish the following remarks:

- The previous results of *Settouti et al.* **[9]** for nucleus recognition were better than ours, but our approach gave the best results according to the Dice coefficient for nucleus and cytoplasm segmentation which demonstrates that our model has a higher sensitivity in detecting the relevant objects.

- Furthermore, the use of the U-Net architecture combined with the pre-trained RESNET-50 encoder allowed for fast learning and segmentation. In contrast to the limitations of other traditional methods that require long processing times **[9]**.

6. Conclusions

This paper reports important results achieved using different models for the encoder part of the U-Net for the automatic recognition of nuclei and cytoplasm regions in cytological images to help experts in medical diagnosis, sometimes even an expert might make a mistake in this so automating the full pipeline. The objective is to automatically detect each object in the image and classify it as a nucleus or a cytoplasm while forming a binary mask to perform the segmentation. Our results were very promising and encouraging, especially by using pre-trained RESNET-50 as the backbone with the loss function adapted to our task, it helped us to addresses the vanishing gradient problem, increasing the segmentation quality and speed.

However, there are further developments in future works to improve the model that performs cell instance segmentation, to distinguish adjacent cells, and to identify individual cells.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the Directorate-General of Scientific Research and Technological Development (Direction Générale de la Recherche Scientifique et du Développement Technologique, DGRSDT, URL: www.dgrsdt.dz, Algeria) for the financial assistance towards this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Al-Dulaimi, K. A. K., Banks, J., Chandran, V., Tomeo-Reyes, I., & Nguyen Thanh, K. (2018). Classification of white blood cell types from microscope images: Techniques and challenges. Microscopy science: Last approaches on educational programs and applied research (Microscopy Book Series, 8), 17-25.
2. Blumenreich, M. S. (1990). The white blood cell and differential count. Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition..
3. Yu, T. C., Chou, W. C., Yeh, C. Y., Yang, C. K., Huang, S. C., Tien, F. M., ... & Chou, S. C. (2019). Automatic Bone Marrow Cell Identification and Classification By Deep Neural Network.
4. Ramakrishna, R. R., Abd Hamid, Z., Zaki, W. M. D. W., Huddin, A. B., & Mathialagan, R. (2020). Stem cell imaging through convolutional neural networks: current issues and future directions in artificial intelligence technology. PeerJ, 8, e10346.
5. Anilkumar, K. K., Manoj, V. J., & Sagi, T. M. (2020). A survey on image segmentation of blood and bone marrow smear images with emphasis to automated detection of Leukemia. Biocybernetics and Biomedical Engineering.

6. Benazzouz, M., Baghli, I., & Chikh, M. A. (2013). Microscopic image segmentation based on pixel classification and dimensionality reduction. International journal of imaging systems and technology, 23(1), 22-28.
7. Baghli, I., Nakib, A., Sellam, E., Benazzouz, M., Chikh, A., & Petit, E. (2014, March). Hybrid framework based on evidence theory for blood cell image segmentation. In Medical Imaging 2014: Biomedical Applications in Molecular, Structural, and Functional Imaging (Vol. 9038, p. 903815). International Society for Optics and Photonics.
8. Benomar, M. L., Chikh, A., Descombes, X., & Benazzouz, M. (2021). Multi-feature-based approach for white blood cells segmentation and classification in peripheral blood and bone marrow images. International Journal of Biomedical Engineering and Technology, 35(3), 223-241.
9. Settouti, N., Bechar, M. E. A., Daho, M. E. H., & Chikh, M. A. (2020). An optimised pixel-based classification approach for automatic white blood cells segmentation. International Journal of Biomedical Engineering and Technology, 32(2), 144-160.
10. Saidi, M., Bechar, M. E. A., Settouti, N., & Chikh, M. A. (2018). Instances selection algorithm by ensemble margin. Journal of Experimental & Theoretical Artificial Intelligence, 30(3), 457-478.
11. Settouti, N., Saidi, M., Bechar, M. E. A., Daho, M. E. H., & Chikh, M. A. (2020). An instance and variable selection approach in pixel-based classification for automatic white blood cells segmentation. Pattern Analysis and Applications, 1-18.
12. Bechar, M. E. A., Settouti, N., Daho, M. E. H., Adel, M., & Chikh, M. A. (2019). Influence of normalization and color features on super-pixel classification: application to cytological image segmentation. Australasian physical & engineering sciences in medicine, 42(2), 427-441.
13. Bechar, M. E. A., Settouti, N., Daho, M. E. H., & CHIKH, M. A. (2018, October). Semi-supervised Super-pixels classification for White Blood Cells segmentation. In 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS) (pp. 1-8). IEEE.
14. Khouani, A., Daho, M. E. H., Mahmoudi, S. A., Chikh, M. A., & Benzineb, B. (2020). Automated recognition of white blood cells using deep learning. Biomedical Engineering Letters, 10(3), 359-367.
15. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
16. Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). PMLR.
17. Iglovikov, V., & Shvets, A. (2018). Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. arXiv preprint arXiv:1801.05746.
18. Chang, S. W., & Liao, S. W. (2019, October). KUnet: microscopy image segmentation with deep unet based convolutional networks. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 3561-3566). IEEE.
19. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
20. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
21. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
22. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.
23. Yakubovskiy, P. Segmentation Models. (2019). https://github.com/qubvel/ segmentation_models.
24. Prechelt, L. (1998). Early stopping-but when?. In Neural Networks: Tricks of the trade (pp. 55-69). Springer, Berlin, Heidelberg.
25. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
26. Yi-de, M., Qing, L., & Zhi-Bai, Q. (2004, October). Automated image segmentation using improved PCNN model based on cross-entropy. In Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004. (pp. 743-746). IEEE.

27. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 240-248). Springer, Cham.