RESEARCH ARTICLE

A Semantic Segmentation of Nucleus and Cytoplasm in Pap-smear Images using Modified U-Net Architecture

Muhammad Arhami[1,*], Fachri Yanuar Rudi F.[2], Hendrawaty[3], and Adriana[4]

[1,2,3]Department of Information Computer Technology, Politeknik Negeri Lhokseumawe, Lhokseumawe 24301, Indonesia
[4]Department of Chemical Engineering, Politeknik Negeri Lhokseumawe, Lhokseumawe 24301, Indonesia

*Corresponding email: muhammad.arhami@pnl.ac.id

Abstract: Pap-smear images can help in the early detection of cervical cancer, but the manual interpretation by a pathologist can be time-consuming and prone to human error. Semantic segmentation of the cell nucleus and cytoplasm plays an essential role in Pap-smear image analysis for automatically detecting cervical cancer. This research proposes a modified U-Net architecture by adding batch normalization to each convolution layer. Batch normalization aims to accelerate the convergence of the weight during training, thus overcoming the vanishing gradient problem. The application of U-Net and batch normalization to pap-smear image segmentation provides good performance results, including accuracy of 91.4 %, specificity of 87.7 %, F1-score of 81.7 %, and precision of 83.7 %. Unfortunately, the sensitivity result obtained is only 79.9 %. The results show that the proposed modification of the U-Net architecture with batch normalization improves the segmentation performance for cervical cancer cells in pap-smear images. However, improvement in architecture is still required to increase the ability to overcome overlapping areas between the nucleus, cytoplasm, and background.

Keywords: semantic segmentation, nucleus, cytoplasm, U-Net architecture, batch normalization

1 Introduction

Cervical cancer is one of the most common cancers in women worldwide. The disease is characterized by the uncontrolled growth of malignant cells in the cervix or cervix area. According to the World Health Organization (WHO), in 2020, 604,000 women in the world are expected to develop cervical cancer, and about 342,000 women will die from the disease [1]. Detecting and diagnosing cervical cancer is very important to increase the patient's chance of recovery and reduce the mortality rate. The pap-smear examination has become a commonly used method for early detection of cervical cancer. Samples of cervix cells are collected and analyzed under a microscope for indications of precancerous or cancerous changes [2]. However, manual interpretation and analysis by pathologists of pap-smear images is usually complex and time-consuming. In addition, there is a risk of human error in identifying and categorizing cervical cancer cells with high accuracy [3]. An automatic diagnosis system is needed to analyze pap-smear images and diagnose cervical cancer quickly and accurately, one of which is image segmentation. Cervical cancer cell segmentation is very important in pap-smear image examination because cervical cancer cells can provide important information about the presence and severity of cervical cancer [4].

Research by Wijaya *et al.* [5] segmented the nucleus and cytoplasm of cells in pap-smear images using the Markov Random Field method. This research only obtained an accuracy value of about 75%, while other evaluation values were not calculated. Other research by Purwono *et al.* [6] segmented cervical cancer cells on CT-scan images using the K-Nearest Neighbors (KNN) method. This research also only obtained an accuracy value between 57-62%, while other evaluation values were not calculated. However, both researchers still used conventional methods. Conventional methods need to improve in distinguishing one object from another, especially in complex images that have many details.

The use of deep learning techniques has grown in recent years. Convolutional Neural Network (CNN) is one such deep learning method that has made significant progress in complex image analysis. A CNN architecture commonly used in complex image analysis is the U-Net. U-Net has the advantage of segmenting and diagnosing diseases accurately [7]. Research by Zhang *et al.* [8] segmented cervical cancer cell images using dilated CNN. This research resulted in an F1-score and precision below 83%, while other evaluation values were not calculated. Another research by Li *et al.* [9] segmented cell nucleus and cytoplasm images using GDLA U-Net. However, the precision and sensitivity obtained for cytoplasmic cells are still below 80%. However, both researches only performed binary segmentation. Semantic segmentation is required to detect cervical cancer cells accurately. Semantic segmentation in cervical cancer involves extracting the nucleus, cytoplasm, and background objects simultaneously rather than just one of the cells.

U-Net architecture is one of the suitable architectures for semantic segmentation as it is a deep network. However, the number of layers in the U-Net architecture can increase the parameters and complexity of the network. A too complex network can hinder the convergence of the weights and cause vanishing gradients [10]. Batch Normalization is a regularization method applied to accelerate convergence and enhance stability during the training process. Batch Normalization works by normalizing the input to each layer in the network [11]. Research by Ju *et al.* [12] conducted cervical cancer CTV image segmentation using the addition of batch normalization to the encoder path on a Dense V-Net architecture. The result obtained is the F1-score value reaches 87.5 %. However, this research only

used 113 CT data and only performed binary segmentation. Another research by Rhee *et al.* [13] also segmented CT scan images of cervical cancer using the addition of batch normalization at the end of each convolution layer. The average F1-score value obtained is 86 %. The data used is quite large, namely 2254 CT data, but only performs binary segmentation.

This research proposes a modification to the U-Net architecture with batch normalization. Batch normalization is added to each convolution layer on the encoder and decoder paths of the U-Net architecture. The addition of batch normalization can reduce the variation of input distribution to the network layers during the training process, thus accelerating weight convergence. The addition of batch normalization to the U-Net architecture is expected to improve the model's performance in performing semantic segmentation with 3 labels (nucleus, cytoplasm, and background) on pap-smear images.

2   Research Method

The workflow in this research is divided into several steps. These steps are data description, pre-processing, training data, testing data, and performance evaluation. The workflow in this research is represented in Figure 1.
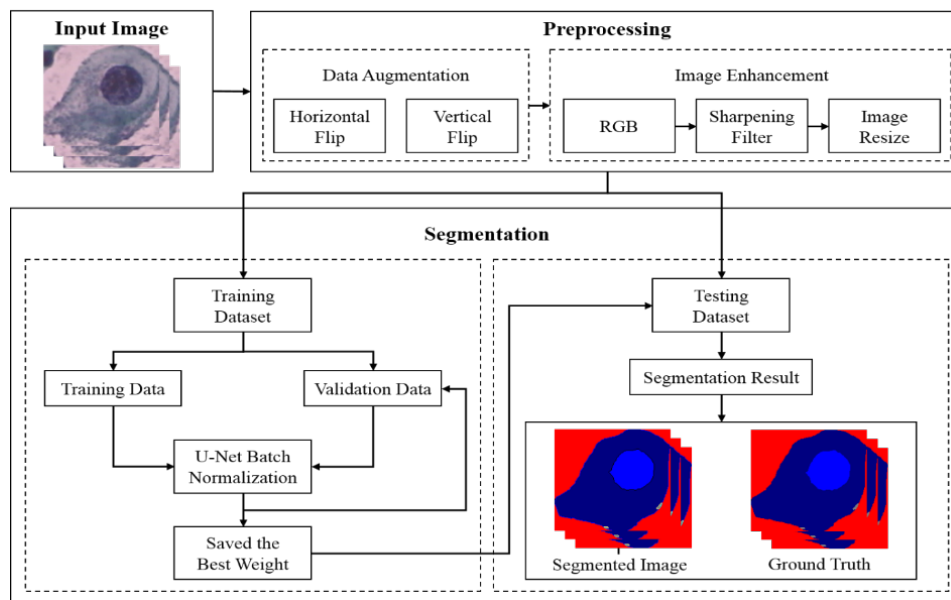

Figure 1: Workflow of the research in cervical cancer cell segmentation.

2.1   Data Description

This research uses the dataset Herlev pap-smear comprising 917 BGR images in .BMP (Bitmap image file) format. This dataset was obtained from Herlev University Hospital at the Department of Pathology and can be accessed through the website [14]. Images of

pap-smears have different dimensions and resolutions. The structure of the nucleus and cytoplasm within the pap-smear image is shown in Figure 2. In Figure 2, it shows that the structural part of the pap-smear image consists of the nucleus (cell nucleus) labeled by the red circle and the cytoplasm (cells surrounding the nucleus) labeled by the blue circle. The structure of the nucleus and cytoplasm is what the ophthalmologist uses as a way to diagnose cervical cancer.
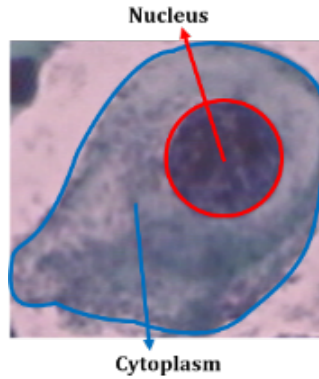

Figure 2: Nucleus and cytoplasm structure in pap-smear image.

2.2   Preprocessing

Preprocessing is the initial image processing process that aims to improve and increase the image quality.

2.2.1   Data augmentation

Data augmentation is a technique used to increase the number of training data. Data augmentation aims to make the model created identified and well-recognized [15]. The data augmentation technique employed in this research is flipping, which involves duplicating the data by flipping the image horizontally or vertically [16].

2.2.2   Image enhancement

Image enhancement aims to remove noise, increase contrast, and preserve all details in the image to prevent any loss of information. Several image quality enhancement techniques used in this research include sharpening filters and image resizing. A sharpening filter is a technique that enhances contrast by sharpening object boundaries and details in the image. This technique is accomplished by increasing the intensity differences between adjacent pixels [17]. Mathematically, the sharpening filter is computed using the Laplacian filter approach using (1) [18].

$$\left(\nabla^2 S\right)(x, y) = \frac{\partial^2 S}{\partial x^2}(x, y) + \frac{\partial^2 S}{\partial y^2}(x, y) \tag{1}$$

Where, ∇^2 is the Laplace operator, $S(x, y)$ is a two-dimensional image function of -axis and -axis. After applying the sharpening filter, the next step is to resize the images to the same dimensions using image resize. Image resize is a method used in the field of image processing that involves changing the pixel size of an image without altering the essential information contained within the image [19].

2.3   Semantic Segmentation

Semantic segmentation is a method within digital image processing that focuses on recognizing and separating image objects at the pixel level. This involves labeling each pixel based on existing categories or classes of objects [20]. Semantic segmentation in cervical cancer involves extracting the nucleus, cytoplasm, and background objects. Some of the operations performed in semantic segmentation include:

2.3.1   Convolutional layer

The convolution layer is the base layer in CNN performing convolution operations on the input images. This layer consists of some filters or kernels that are shifted gradually on the input image to generate feature maps. The convolutional layer learns the visual features representation of the input image through a convolution process with customized filters or kernels. The convolution calculation process in the convolutional layer is obtained using (2) [21].

$$a_{i,j} = \left(\sum_{u=0}^{n-1} \sum_{v=0}^{n-1} d_{u+i,v+j} \times k_{u+1,v+1} \right) + b_q \tag{2}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$, a_{ij} represents the entry of the input matrix resulting from the convolution process at the i-th row and the j-th column, $d_{u+i,v+j}$ represents the entry of the input matrix at the $u + i$-th row and $v + j$-th column, $k_{u+1,v+1}$ represents the entry of the kernel matrix at the $u + 1$-th row and $v + 1$-th column and b_q is the biar for the q-th kernel.

2.3.2   Batch normalization

Batch normalization is a normalization process performed on each layer within a CNN network, aiming to improve accuracy and time efficiency during the training process. The batch normalization process is carried out by calculating the mean value (μ_j) and variance (σ_j^2) for each mini-batch using (3) and (3) [21].

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} a_{ij} \tag{3}$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} (a_{ij} - \mu_j)^2 \tag{4}$$

where, j represents the count of columns within the mini-batch, m represents the quantity of data present in one mini-batch, and a_{ij} represents the entry within the input matrix at

the i-th row and j-th column. Furthermore, the entry of the input matrix (a_{ij}) is normalized using (5).

$$\hat{a}_{ij} = \frac{a_{ij} - \mu_j}{\sqrt{\sigma_j^2 + \varepsilon}} \tag{5}$$

where, $\hat{a}_{ij}$ is the entry of the normalized matrix, and is the smallest constant value.

2.3.3   Activation function

The activation function serves as a non-linear function utilized for the purpose of introducing non-linearity and complex mapping capabilities in a CNN network. The activation function does not change the dimensions of the feature maps but only alters the values of the input feature maps [22]. The activation functions used in this research are rectified linear unit (ReLU) and softmax. The ReLU activation function is a non-linear function that assigns a value of 0 to all negative pixel values within an image. The calculation of the ReLU activation function is obtained using (6) [22].

$$r(\hat{a}_{ij}) = \max(0, \hat{a}_{ij}) = \begin{cases} \hat{a}_{ij} & \text{if}, \hat{a}_{ij} \geq 0 \\ 0 & \text{if}, \hat{a}_{ij} < 0 \end{cases} \tag{6}$$

where, $\hat{a}_{ij}$ is the input value of the image and $r(\hat{a}_{ij})$ is the output result of the ReLU. The softmax activation function is a mathematical function utilized to compute the probabilities for each predicted label, where the probabilities are exponential probabilities normalized from the class observations. The softmax activation function is obtained using (7) [23].

$$s(t)_j = \frac{e^{t_j}}{\sum_{k=1}^{K} e^{t_k}} \tag{7}$$

for $k = 1, \ldots, K$ where K represents the quantity of classes and t_j represents the entry of the input matrix.

2.3.4   Max pooling layer

The max pooling layer is one of the types of pooling layers that diminishes the dimensionality of the feature maps produced by the preceding layer. It achieves this by extracting the patch from the convolutional feature maps and selecting the highest value in each segment to undergo shifting [24].

2.3.5   Transposed convolution

A Transposed convolution is a convolutional layer used to increase the dimensionality of the input by inserting zeros between adjacent elements. This layer performs the inverse operation of a regular convolutional layer [25].

2.3.6   Concatenate layer

The concatenate layer is a layer in a CNN network used to combine the outputs from multiple preceding layers into one. In this layer, the concatenation is done horizontally by

combining information from different layers and features obtained from different levels of hierarchy in the network [26].

2.3.7   Loss function

The loss function is a metric utilized during the process of training a model to assess the discrepancy or gap between the expected (ground truth) values and the model's predicted values. In semantic segmentation, the loss function commonly used for multiclass labels or labels with more than two object classes is categorical cross-entropy. The categorical cross-entropy value is obtained using (8) [27].

$$L(y, s) = -\sum_{i=1}^{m} y_i \times \log s_i \qquad (8)$$

where, m represents the number of rows within the resultant output matrix, s_i is the entry of the predicted segmentation output matrix at the i-th row, y_i represents the entry of the ground truth matrix at the i-th row, and L is the value of the resulting categorical cross-entropy.

2.4   Modified Architecture

The semantic segmentation process of the nucleus and cytoplasm is performed by applying the U-Net architecture with the addition of batch normalization into every convolutional operation. The addition of batch normalization aims to enhance stability and training speed, as well as help in overcoming the vanishing gradient problem. The modification of the architecture proposed in this research for performing semantic segmentation is shown in Figure 3. It shows the modified U-Net architecture consisting of two paths: the left side containing the encoder path and the right side containing the decoder path. The encoder path includes a convolution block, batch normalization, ReLU activation, and max pooling. Meanwhile, the decoder path consists of a convolution block, transposed convolution, and softmax. The encoder path begins with a convolution operation using a 3×3 kernel and filters. This convolution process is performed concurrently alongside the ReLU activation function. Next, the resulting feature maps from the convolution process will undergo a batch normalization process to be normalized. Then, a max pooling operation with a size of 2×2 is performed to reduce the dimension of the feature maps. In the encoder path, there are four convolution blocks, where each block doubles the number of feature maps using filters of sizes 64, 128, and 256 respectively. This is followed by a fifth block that serves as a bridge between the path of the encoder and the decoder. It involves identical procedures to those employed in the initial block but without the need for subsequent max pooling. Next, the decoder path begins with a transposed convolution operation of size 2×2, performed simultaneously with the concatenate operation between the feature maps from the encoder path and the feature maps from the decoder path. This step aims to restore the dimensionality of the feature maps to their original size. Then, the decoder path continues with the same process as the first block in the encoder path, without using max pooling. In the decoder path, there are four convolution blocks, where the count of feature maps in each block is divided by two until it returns to the original count of feature maps. The final step in the decoder path is a convolution process with a 1×1 kernel, performed simulta-

neously with the softmax activation function. This process aims to generate an image that has undergone segmentation by obtaining probabilities for each object class.
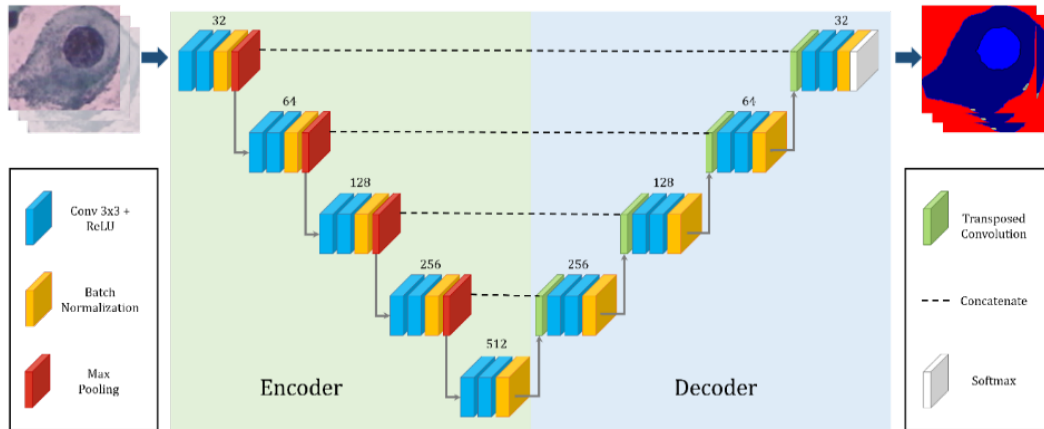

Figure 3: Modified U-Net architecture with batch normalization.

2.5 Evaluation

In this research, a performance evaluation is carried out on the results of image enhancement that has been improved using the sharpening filter method. This performance evaluation uses the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metrics (SSIM) metrics. Furthermore, in the semantic segmentation of the nucleus and cytoplasm, each pixel is grouped into three classes: nucleus cells, cytoplasm cells, and background. Evaluation of the model's performance in the semantic segmentation process is done using the confusion matrix. These results of the methods used in semantic segmentation provide insight into the U-Net Batch Normalization architecture's performance in accurately segmenting the nucleus and cytoplasm. In this research, the performance evaluation metrics used include accuracy, sensitivity, specificity, F1-score, and precision.

3 Result

3.1 Preprocessing

Data augmentation can improve the amount of training data without losing semantic information and help reduce bias in the data. In this research, horizontal and vertical flipping techniques were used in the data augmentation process. The data augmentation process on pap-smear images is shown in Figure 4. It shows that in the horizontal reversal technique, the image is rotated horizontally, while in the vertical reversal technique, the image is rotated vertically. This creates a new variation in the dataset by changing the direction or orientation of the images. The original Herlev dataset consists of 917 images. Through the data augmentation process, the total amount of data increased to 2,751 with each addition

of data from vertical flipping. Furthermore, the augmented images undergo a process of image quality enhancement, where the flow of the image quality enhancement process is shown in Figure 5. It shows that the result of data augmentation is used as an input image of type BGR. Then, the BGR image is converted to an RGB image. Images of RGB are then subjected to contrast enhancement using the sharpening filter method. The goal is to make the nucleus and cytoplasm structures appear clearer and sharper. Furthermore, the image undergoes an image resize process, changing its size to 256×256 pixels.

Image resize is a technique used to change the pixel size in an image without altering the important information contained in it. In this research, quantitative image quality measurements are performed by comparing the PSNR and SSIM values between the original image and the preprocessed image.
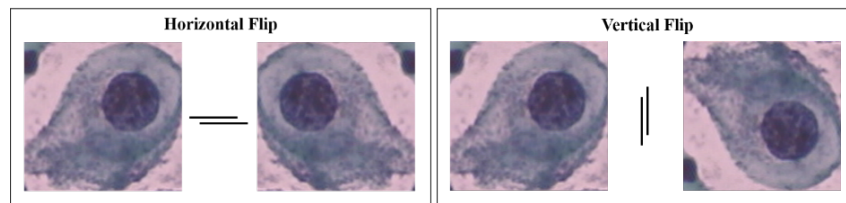
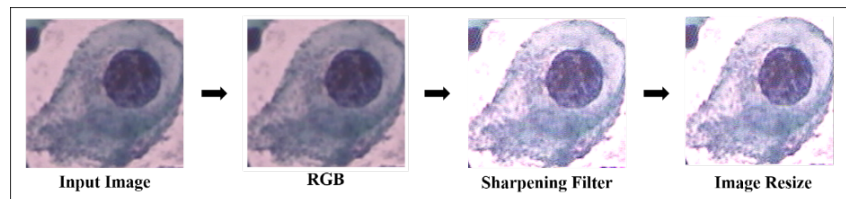Figure 4: Data augmentation using horizontal flipping and vertical flipping.

Figure 5: Flow of the image quality enhancement process.

The measurement results are presented in a comparison graph as shown in Figure 6. It, shows that the performance evaluation results using the sharpening filter method show the average PSNR and SSIM values that have approached or reached a number that is considered good. The PSNR value graph in Figure 6(a) shows the average PSNR value is 42.887. The PSNR value is used to measure the level of noise or distortion in the image after the preprocessing process. If the value of PSNR is higher, then the noise level in the enhanced image is lower. Meanwhile, Figure 6(b) shows a graph of the SSIM value with an average value is 0.908. A high SSIM value indicates good structural similarity between the enhanced image and the ground truth. Thus, it can be said that the image quality after enhancement is good.

3.2   Training Data

The training data process was performed using the preprocessed results, totaling 2,751 data, then split into 80 % training data and 20 % testing data. This resulted in approximately 2,200 training data randomly split. Furthermore, this training data was further divided
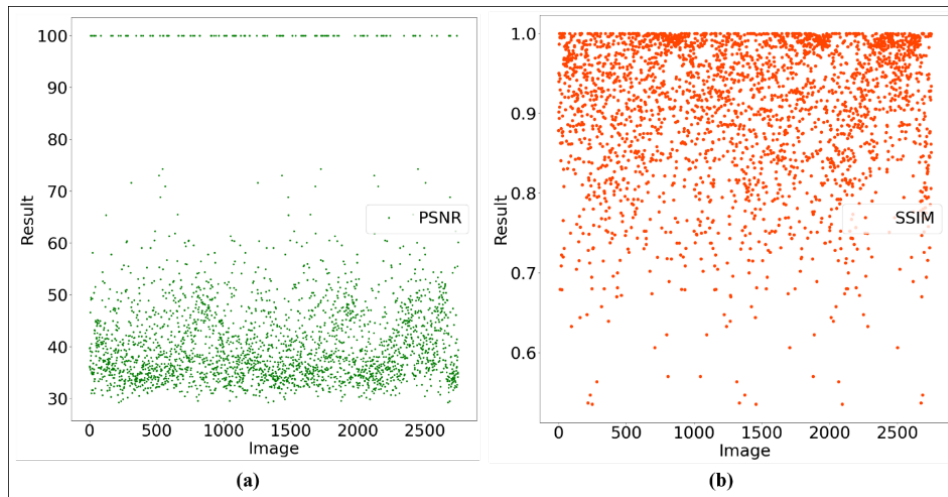
Figure 6: Comparison graph of values between the original image and the preprocessed image: (a) PSNR and (b) SSIM.

using a split percentage of 70 % (1,540) in training data and 30 % (660) in validation data. At this stage, the modified architecture of U-Net as shown in Figure 3 is used for the training process. The training process in this research utilizes 30 epochs. The results of the training process yield the values of accuracy and loss in the training data as well as the values of val_accuracy and val_loss for the validation data. The graph of results obtained during the training process with the modified U-Net architecture model is shown in Figure 7.
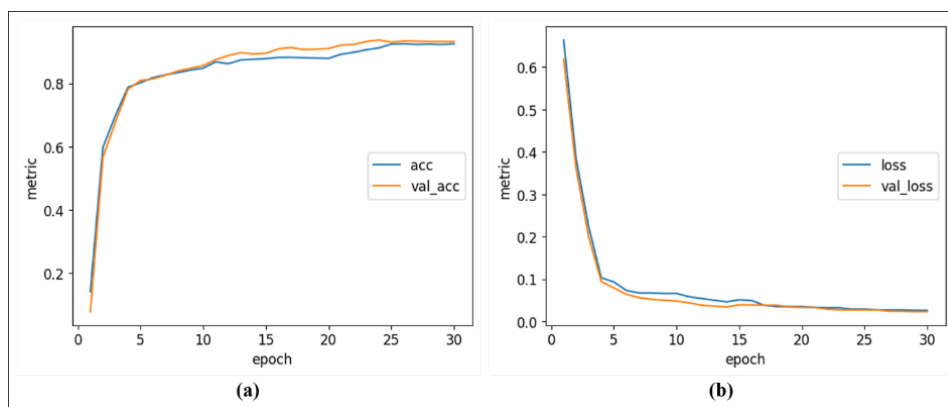
Figure 7: Graphs obtained during the training process: (a) accuracy and (b) loss.

Figure 7(a) represents the graph of accuracy for training data and validation data using the modified U-Net architecture. In the training data, the accuracy graph shows an increase in each epoch. Starting from 14 % in the first epoch, this number continues to increase until reaching 92 %. The same thing can be observed in the accuracy graph for the validation

data, which also increases with each epoch. Starting from 7 % in the first epoch, this number continues to increase until it reaches 93 %. Meanwhile, Figure 7(b) represents the graph of loss for training data and validation data using the modified U-Net architecture. For the training data, the loss graph shows a decrease at each epoch. In the first epoch, the loss value is 66 %, then it gradually decreases to near 0 %. Similarly, for the validation data, the loss graph also shows a decrease in each epoch. In the first epoch, the loss value is 61 %, then it continues to decrease toward 0 %. Smaller loss values indicate a smaller error rate in the model's learning of patterns.

The results of the graphs indicate that the used model does not experience overfitting and is capable of recognizing and learning patterns in the trained data. Based on Figure 7(a) and Figure 7(b), the performance of the modified U-Net architecture model is good in nucleus and cytoplasm segmentation, as indicated by an accuracy above 90 % and a loss value approaching 0 %.
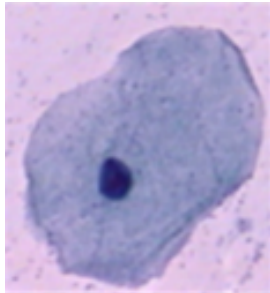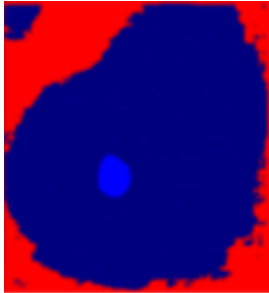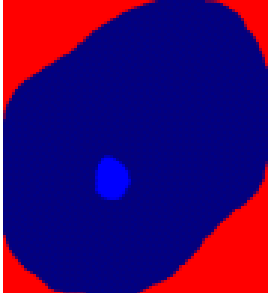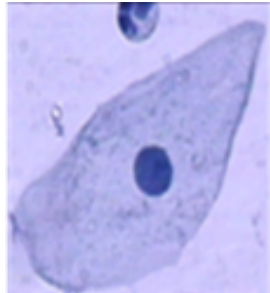
3.3   Testing

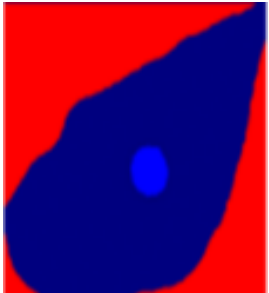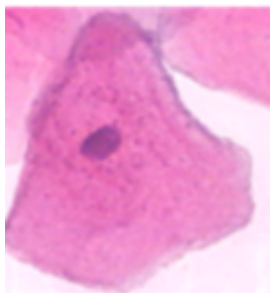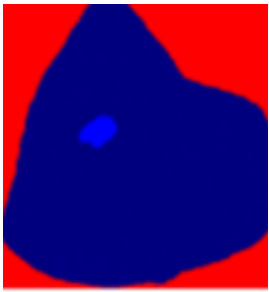The testing process is a step to test the model from the results of the training process using new data that has never been learned by the model before. The testing data consists of 551 data obtained from split data. At this stage, semantic segmentation predictions are performed for the nucleus and cytoplasm are performed, and evaluate the accuracy of the model.

Several comparisons between the original image, segmentation result, and ground truth are shown in Table 1. It shows the comparison between the segmented image and the ground truth. The structure of the pap-smear image consists of the nucleus (cell nucleus) labeled in light blue, the cytoplasm (cells surrounding the nucleus) labeled in dark blue, and the background labeled in red. Seen in Table 1, the segmentation results performed using the modified U-Net architecture with Batch Normalization have shown similarity with the ground truth. However, the segmentation results of the nucleus area are still not fully predictable. In addition, in some results, there are still points in the background that are incorrectly predicted as cytoplasm.

The performance evaluation metrics used for semantic segmentation of nucleus and cytoplasm cells include accuracy, sensitivity, specificity, F1-score, and precision. Accuracy is used to measure the extent to which the segmentation model can correctly identify between nucleus and cytoplasm cells in pap-smear images. Sensitivity is used to measure the ability of the model to correctly identify cancer cells, including both nucleus and cytoplasm cells. Specificity is used to measure the ability of the model to correctly identify the background. Precision is used to measure how precise the model is in identifying nucleus and cytoplasm cells. F1-score is used for the harmonic mean between precision and sensitivity. A comparison of the obtained semantic segmentation performance evaluation results with other studies is shown in Table 2.

Table 2 shows the comparison of research results using the same dataset for pap-smear image segmentation. It is observed that the semantic segmentation method proposed in this research achieved the highest values in terms of accuracy, F1-score, and specificity compared to previous studies. Specifically, the accuracy is 91.48 %, F1-score is 81.7 %, and specificity is 87.7 %. However, another study by [30] obtained the highest precision, and a study by [31] obtained the highest sensitivity, although these two studies only calculated three evaluation performance values. Compared to other studies, it can be seen

Table 1: Comparisons of Original Image, Segmentation Result, and Ground Truth

No	Pap-smear Dataset		
	Original Image	**Segmentation Result**	**Ground Truth**
1			
2			
3			

that these studies only measured 2 to 4 evaluation performance metrics. According to the comparison, it is concluded that the proposed method has provided optimal performance in semantic segmentation.

4   Discussion

In the process of nucleus and cytoplasm segmentation using the U-Net architecture, each pixel is grouped into three different classes. Class 0 is used for the cytoplasm label, class 1 for the nucleus label, and class 2 for the background label. A comparison of the performance evaluation of each label is shown in Figure 8. This figure shows the performance

Table 2: Evaluation result comparison

Method	Acc (%)	F1 (%)	Sen (%)	Spe (%)	Pre (%)
SPFNet [28]	-	-	-	-	78.40
Bi-Path CNN [29]	77.00	69.00	72.00	-	60.00
Deep CNN Without Pre-processing [30]	-	77.00	70.00	-	90.00
U-Net [31]	-	77.00	94.74	75.00	-
Proposed method	91.48	81.70	79.90	87.70	83.70

evaluation results of each label, where it can be seen that class 1 has the highest accuracy and specificity values compared to other classes. Meanwhile, class 0 obtains higher F1-score, sensitivity, and precision values compared to the other classes. The performance evaluation per label indicates that the accuracy obtained for all classes is very good, with results that are close to the ground truth. The F1-score shows the model's excellent performance in segmenting the nucleus and cytoplasm. Sensitivity shows that the models have a high ability to detect nucleus and cytoplasm objects, with higher values than the background. Specificity shows that the models have a high ability to detect objects other than the nucleus and background, although the specificity for the cytoplasm is slightly lower. Precision shows that the model can accurately identify boundaries for the cytoplasm, nucleus, and background, thereby reducing errors in classifying adjacent pixels for each object.
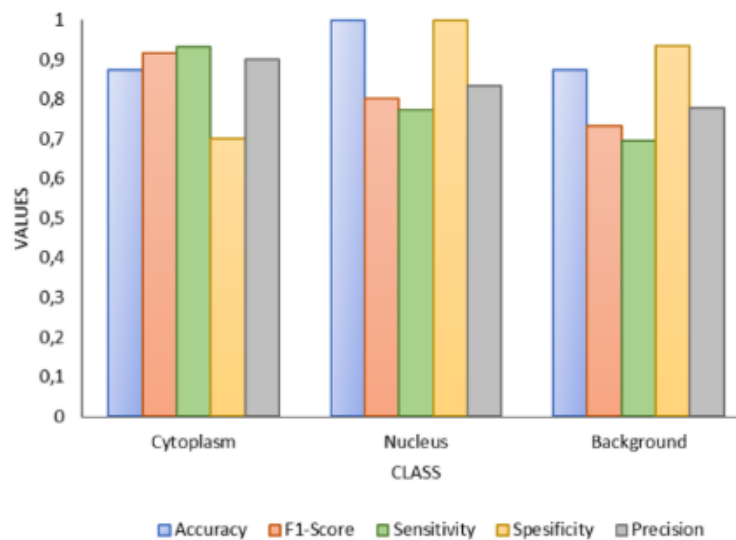

Figure 8: Comparison of performance evaluation results per label.

5   Conclusion

Based on previous research, the use of modified U-Net architecture for nucleus and cytoplasm segmentation has been proven effective in predicting the pixels representing the nucleus and cytoplasm from the given image data. This research modified the U-Net architecture by adding batch normalization in the semantic segmentation process but did not involve the classification process. Therefore, the future research of this research will focus on classification for cervical cancer detection based on the segmentation results obtained. Performance evaluation shows that the modified U-Net architecture has provided good segmentation results. The problem of network complexity and vanishing gradient during the training process was successfully overcome by the addition of batch normalization to the basic U-Net architecture. This led to accurate segmentation predictions based on the evaluation values obtained.

Acknowledgments

References

[1] WHO, "Who guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention, second edition: use of mrna tests for human papillomavirus (hpv)," tech. rep., 2021.

[2] K. Raimond, G. B. Rao, S. Juliet, S. R. G. Tamilarasi, P. S. Evangelin, and L. Mathew, "An emerging paradigms on cervical cancer screening methods and devices for clinical trails," *Frontiers in Public Health*, vol. 10, p. 1030304, Oct. 2022.

[3] L. Allahqoli, A. S. Laganà, A. Mazidimoradi, H. Salehiniya, V. Günther, V. Chiantera, S. Karimi Goghari, M. M. Ghiasvand, A. Rahmani, Z. Momenimovahed, and I. Alkatout, "Diagnosis of Cervical Cancer and Pre-Cancerous Lesions by Artificial Intelligence: A Systematic Review," *Diagnostics*, vol. 12, p. 2771, Nov. 2022.

[4] I. T. Hoque, N. Ibtehaz, S. Chakravarty, M. S. Rahman, and M. S. Rahman, "A contour property based approach to segment nuclei in cervical cytology images," *BMC Medical Imaging*, vol. 21, p. 15, Dec. 2021.

[5] R. S. D. Wijaya, Adiwijaya, Andriyan B Suksmono, and Tati Lr Mengko, "Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, pp. 139–147, Feb. 2021.

[6] R. R. P. A. Purwono, E. Purwanti, and R. Rulaningtyas, "Segmentation of cervical cancer CT-scan images using K-nearest neighbors method," (Surabaya, Indonesia), p. 040009, 2020.

[7] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical Image Segmentation based on U-Net: A Review," *Journal of Imaging Science and Technology*, vol. 64, pp. 020508–1–020508–12, Mar. 2020.

[8] M. Zhang, J. Zhao, X. Li, L. Zhang, and Q. Li, "ASCNET: Adaptive-Scale Convolutional Neural Networks for Multi-Scale Feature Learning," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, (Iowa City, IA, USA), pp. 144–148, IEEE, Apr. 2020.

[9] G. Li, C. Sun, C. Xu, Y. Zheng, and K. Wang, "Cervical Cell Segmentation Method Based on Global Dependency and Local Attention," *Applied Sciences*, vol. 12, p. 7742, Aug. 2022.

[10] S. Ma, J. Tang, and F. Guo, "Multi-Task Deep Supervision on Attention R2U-Net for Brain Tumor Segmentation," *Frontiers in Oncology*, vol. 11, Sept. 2021.

[11] Y. Cui, J. Huang, and D. Wu, "Optimize TSK Fuzzy Systems for Classification Problems: Mini-Batch Gradient Descent with Uniform Regularization and Batch Normalization," *IEEE Transactions on Fuzzy Systems*, vol. 28, pp. 3065–3075, Dec. 2020. arXiv:1908.00636 [cs, stat].

[12] Z. Ju, W. Guo, S. Gu, J. Zhou, W. Yang, X. Cong, X. Dai, H. Quan, J. Liu, B. Qu, and G. Liu, "CT based automatic clinical target volume delineation using a dense-fully connected convolution network for cervical Cancer radiation therapy," *BMC Cancer*, vol. 21, p. 243, Dec. 2021.

[13] D. J. Rhee, A. Jhingran, B. Rigaud, T. Netherton, C. E. Cardenas, L. Zhang, S. Vedam, S. Kry, K. K. Brock, W. Shaw, F. O'Reilly, J. Parkes, H. Burger, N. Fakie, C. Trauernicht, H. Simonds, and L. E. Court, "Automatic contouring system for cervical cancer using convolutional neural networks," *Medical Physics*, vol. 47, pp. 5648–5658, Nov. 2020.

[14] "Downloads – MDE-Lab."

[15] A. F. M. S. Uddin, M. S. Monira, W. Shin, T. Chung, and S.-H. Bae, "SaliencyMix: A Saliency Guided Data Augmentation Strategy for Better Regularization," July 2021. arXiv:2006.01791 [cs, stat].

[16] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *Journal of Imaging*, vol. 9, p. 46, Feb. 2023.

[17] H. A. Siregar, "Implementasi metode sharpening untuk memperbaiki kualitas citra," *Pelita Informatika: Informasi dan Informatika*, vol. 8, no. 1, pp. 5–8, 2019.

[18] I. Ummah and N. Yannuansa, "Analisis pendeteksian tepi objek pada pengolahan citra," *Prosiding Seminar Nasional Sains, Teknologi, Ekonomi, Pendidikan dan Keagamaan (SAINSTEKNOPAK)*, vol. 4, Dec. 2020.

[19] N. A. Deepak, G. Savitha, D. J. Deepak, and P. K. Supraj, "Retinopathy Based Multistage Classification of Diabeties:," (Bangalore, India), 2021.

[20] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[21] "BVU-Net: A U-Net Modification by VGG-Batch Normalization for Retinal Blood Vessel Segmentation," *International Journal of Intelligent Engineering and Systems*, vol. 15, pp. 303–314, Dec. 2022.

[22] T. A. Soomro, A. J. Afifi, L. Zheng, S. Soomro, J. Gao, O. Hellwich, and M. Paul, "Deep Learning Models for Retinal Blood Vessels Segmentation: A Review," *IEEE Access*, vol. 7, pp. 71696–71717, 2019.

[23] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *International Journal of Engineering Applied Sciences and Technology*, vol. 04, pp. 310–316, May 2020.

[24] T. Naraloka, L. I. Kesuma, A. Sukmawati, and M. Cristianti, "Arsitektur U-Net pada Segmentasi Citra Hati sebagai Deteksi Dini Kanker Liver," *Techno.Com*, vol. 21, pp. 753–764, Nov. 2022.

[25] D. Im, D. Han, S. Choi, S. Kang, and H.-J. Yoo, "DT-CNN: An Energy-Efficient Dilated and Transposed Convolutional Neural Network Processor for Region of Interest Based Image Segmentation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, pp. 3471–3483, Oct. 2020.

[26] M. Gan and C. Wang, "Dual-Stage U-Shape Convolutional Network for Esophageal Tissue Segmentation in OCT Images," *IEEE Access*, vol. 8, pp. 215020–215032, 2020.

[27] Y. Ho and S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.

[28] M. Xia, G. Zhang, C. Mu, B. Guan, and M. Wang, "Cervical Cancer Cell Detection Based on Deep Convolutional Neural Network," in *2020 39th Chinese Control Conference (CCC)*, (Shenyang, China), pp. 6527–6532, IEEE, July 2020.

[29] A. Desiani, Erwin, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, "Bi-path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap-smear Images. | IAENG International Journal of Computer Science | EBSCOhost," Sept. 2021.

[30] S. Gautam, A. Bhavsar, A. K. Sao, and H. K.K., "CNN based segmentation of nuclei in PAP-smear images with selective pre-processing," in *Medical Imaging 2018: Digital Pathology* (M. N. Gurcan and J. E. Tomaszewski, eds.), (Houston, United States), p. 32, SPIE, Mar. 2018.

[31] N. Nazir, A. Sarwar, B. S. Saini, and R. Shams, "A Robust Deep Learning Approach for Accurate Segmentation of Cytoplasm and Nucleus in Noisy Pap Smear Images," *Computation*, vol. 11, p. 195, Oct. 2023.