Combination of Image Enhancement and U-Net Architecture for Cervical Cell Semantic Segmentation

Rudiansyah1)*, Lemi Iryani2), Lucky Indra Kesuma3), Puspa Sari4), Agung Alamsyah5)

1,2,3) Department of Information Systems, Faculty of Computer Science, Sjakhyakirti University, Indonesia
4,5) Department of Mathematics, Faculty of Mathematics and Natural Sciences, Sriwijaya University, Indonesia

*Corresponding Email: rudiansyah@unisti.ac.id

Abstract

Cervical cancer is the second leading cause of death in women and ranks fourth as a disease that occurs in women worldwide. Cervical cancer is a disease that is difficult to detect and can be detected when it is in an advanced stage. This requires early prevention by carrying out a pap-smear examination. Pap-smear examination manually requires a relatively long time, so a tool is needed by segmentation. Segmentation is image processing by performing perfection between the intended object and the background. One of the CNN methods commonly used in medical image segmentation is the U-Net architecture. Segmentation in this study was carried out on the nucleus and cytoplasm of the Herlev dataset using the U-Net architecture combined with data augmentation and image enhancement. In the learning process, this research resulted in a fairly high IoU value of 78% and an RMSE close to 20%. The results of this study also yielded an accuracy value of 89%, with an average precision, recall and F1 score of 89%, 89% and 88.67%, respectively. This shows that the combination of the CNN U-Net architecture with image quality improvement and data augmentation is quite good at segmenting cervical cells for the nucleus and cytoplasm.

Keywords: Cervical Cancer, Segmentation, CNN, U-Net, Image Enhancement

I. INTRODUCTION

Cervical cancer is the second leading cause of death in women and ranks fourth among cancers that occur in women worldwide (Vu et al., 2018). Based on WHO data in 2018, it is estimated that 570,000 women were diagnosed with cervical cancer, with 311,000 of them dying. The results of the World Cancer Observation (Globocan) in Indonesia stated that cervical cancer was in second place with the addition of more than 36,000 new cases or around 9.2% of the total cancer cases (Maryam & Arino, 2022). Cervical cancer or cervical cancer occurs in the cervical area, which is the entrance between the uterus and the vagina, which is caused by the HPV virus (Esmailizadeh & Nasirzadeh, 2023). Cervical cancer is difficult to detect and can be detected at an advanced stage (Youneszade et al., 2023). One of the efforts to prevent cervical cancer by anticipating it early is the pap smear examination (Somapawong et al., 2019).

A pap smear image is an observational image of a single cell under a microscope that has been processed using a staining method appropriate to the nuclear and cytoplasmic areas (William et al., 2018). However, manual Pap smear examination has several disadvantages, namely that it takes a relatively long time and the opportunity for errors during analysis is large because it is subjective. To reduce the risk of these errors, we need a tool to detect cervical cells by using segmentation. Image segmentation is an image processing process that aims to separate object areas from background areas so that objects can be easily analyzed to recognize objects that involve a lot of visuals (Ghosh et al., 2019). A pap smear image consists of 3 parts, namely background, cytoplasm, and nucleus. Segmentation of the three parts can be done with semantic segmentation.

In performing segmentation, the image used must be of good quality to produce good and accurate segmentation results (Almirat et al., 2019). If the image has sensitive contrast, is blurry, not clear, has spots, and there are parts of detail that are less clear, it is necessary to improve the image quality to produce better image quality (Kwaja Kusuma & Kusumadewi, 2020). Improving image quality is the initial stage of image processing. The initial stage in improving image quality is to increase the contrast.

Contrast enhancement can help in increasing the lighting of an image that has low contrast (Kaur et al., 2018). There are various methods of increasing contrast, namely Contrast Limited Adaptive Histogram Equalization (CLAHE) and Adaptive Gamma Correction. CLAHE is a useful method for increasing image contrast and overcoming color unevenness (Al-hazaimh et al., 2022; Bataineh & Almotairi, 2021). CLAHE is very useful in increasing contrast in dark images (Rudiansyah et al., 2023). CLAHE has the advantage of being able to suppress noise in the same area and increase the contrast and edges of objects (W. Li et al., 2019). In addition to having advantages, the disadvantage of using CLAHE contrast enhancement is that there is image distortion due to excessive enhancement so important features of the image cannot be extracted properly (Awasthi et al., 2020). The disadvantages of CLAHE can be overcome by the Adaptive Gamma Correction method. Gamma Correction has the advantage of increasing the intensity of lighting in an image (Xu et al., 2009). Another step in improving image quality is reducing noise. One method that can be used to reduce noise is to use the Median Filter (Kesuma et al., 2022). The median filter is a non-linear method that can be used to remove noise from an image by replacing each pixel value with the median of that pixel (Win et al., 2011).

Various methods have been developed for segmentation, one of the segmentation methods that can be used with large datasets is Convolutional Neural Network (CNN). The advantage of the CNN method is that it can identify and select features in images in the convolution process (Jogin et al., 2018). CNN has various architectures that can be used to perform segmentation, one of which is the U-Net architecture. The U-net architecture is a medical image architecture that is accurate in segmentation and can increase the accuracy of disease diagnosis (Yin et al., 2022). The U-Net architecture has two paths, namely encoder, and decoder. The encoder process is used to reduce the size of the input matrix, while the decoder path returns the matrix size to its original size by minimizing the number of feature maps so that the image can be segmented properly (Naraloka et al., 2022).

Research on cervical cell segmentation was conducted by Desiani (Desiani et al., 2021) using the U-Net architecture to segment the cervix into 3 parts, namely background, nucleus, and cytoplasm by combining three image improvement methods, namely normalization, CLAHE, and adaptive gamma correction. However, the study (Desiani et al., 2021) did not produce good enough results, with accuracy, sensitivity, specificity, and f1-score below 80%. Another study was conducted by Zhao (Zhao et al., 2019) using the U-Net architecture in segmenting 3 parts, namely background, nucleus, and cytoplasm to produce precision, recall, and f1-score respectively, namely 89%, 87%, and 88%. Research (Zhao et al., 2019) has not calculated accuracy and there is no use of image enhancement in the preprocessing process. Other research was conducted by Li (G. Li et al., 2022) using the U-Net method combined with the GDLA (Global Dependency
and Local Attention) method which resulted in precision and recall of 88.8% and 93.6%. The study (G. Li et al., 2022) did not use image enhancement and accuracy measurement.

This study aims to combine the U-Net architecture for the image improvement process, namely Adaptive Gamma Correction, CLAHE, and Median Filter to segment cervical cells on nuclear and cytoplasmic features. This study aims to assist the medical team in the early detection of cervical cancer to minimize the number of cervical cancer sufferers. The success rate of this study in segmenting cervical cell images was measured by calculating the respective values of accuracy, precision, recall, f1-score, and IOU.

II. RESEARCH METHOD

This research consists of several stages including data preparation, data pre-processing, and image segmentation using U-Net. Data preprocessing consists of data augmentation, image enhancement, and one-hot encoding on labels. Completely, the research stages are shown in the research method in figure 1.

A. Data Preparation

In this research, the Herlev dataset was used which consisted of images and labels of cervical cancer cells to extract the nucleus and cytoplasm which can be accessed on the online site http://mde-lab.aegean.gr/index.php/downloads. The Herlev dataset consists of 917 images with an average size of 150 × 140 pixels. The dataset has also been given ground truth as comparison material when performing semantic segmentation. The segmentation labels used in this research were 0 indicating background, 1 indicating nucleus, and 2 indicating cytoplasm. An example of an image along with a label from Herlev dataset is shown in figure 2.
Based on the figure 2, for class 0 shows the cytoplasm, it is shown in navy blue, class 1 shows the nucleus which is shown in light blue, and class 2, shows the background, it is shown in red.

**B. Data Preprocessing**

Data preprocessing used in this research is in the form of data augmentation, image enhancement, and one-hot encoding.

1) **Data Augmentation**

Data augmentation is a technique in machine learning that purposed to increase the amount and variety of training data. Augmentation produces new data by carrying out transformations on the original data (20). Augmentation data can help improve model performance and prevent overfitting by giving the model more variety in the data training process (Shorten & Khoshgoftaar, 2019). The augmentation method used in this research is:

2) **Flipping Image**

One of the geometric data augmentations is flipping. The flipping used in this research is a horizontal flip and a vertical flip. Horizontal flip is a form of transformation by reflecting the Y axis, while vertical flip is a form of geometric transformation by reflecting the X axis (Zaelani & Miftahuddin, 2022).

3) **Sharpening Kernel**

Kernel sharpening is a form of transformation that sharpens image colors (Gazali et al., 2012). In the image enhancement process, a kernel is needed which is obtained from the image data. This kernel is a matrix whose value can be changed for image improvement, for example, the sharpen kernel. The purpose of sharpening is to increase the detail and sharpness of the image so that it looks clearer and more detailed. Filter on the kernel can be done with equation 1 (Geum et al., 2020).

\[
g(t) = \omega f(t) = \sum_{\tau=0}^{t} f(\tau) g(t - \tau)
\]

Where \(g(t)\) is the filtered image function, \(f(t)\) is the original image function and \(\omega\) is the filter kernel. Each element in the filter kernel will be affected by the interval \(0 \leq \tau \leq t\).

4) **Gaussian Filter**

The Gaussian filter is a type of transformation that blurs the sharp corners of an image to make it smoother (23). Gaussian filters work by dampening the high-frequency components in the image, resulting in a smoother appearance. The Gaussian filter can be represented by Equation 2 (Yuwono, 2015).

\[
G(x, y) = c \cdot \exp \left( -\frac{(x - u)^2 + (y - v)^2}{2\sigma^2} \right)
\]

Where \(G(x, y)\) is an element of the filtered kernel matrix, \(c\) is a constant, \(\sigma^2\) is the variance. At least, \(u\) and \(v\) are the central indices of the gaussian kernel matrix.

5) **Image Enhancement**

The image quality in the Herlev dataset is characterized by low contrast and inconsistent colors. Therefore, in this research, contrast enhancement is applied to improve image quality. The image enhancement techniques employed in this research are adaptive gamma correction, CLAHE, and median filtering.

6) **Adaptive Gamma Correction**

Adaptive Gamma Correction is a non-linear operation to increase image contrast using a power transformation approach (Singh et al., 2010). Adaptive Gamma Correction is defined in equation (3).
Where $A$ is the brightness value of the resulting image, $R$ is the brightness value of the original image, $\gamma$ is the level of image brightness. The value $\gamma < 1$ indicates the new image is lighter than the original image, while $\gamma > 1$ indicates the new image is darker than the original image. $T(R)$ is the adaptive gamma correction transformation function of $R$ (Desiani et al., 2021).

7) **CLAHE**

CLAHE is used to increase image contrast and overcome uneven colors (Maria et al., 2018). CLAHE aligns the 3 RGB color components which are calculated using the clip limit constraint from the histogram contained in equation (4).

$$
\beta = \frac{M}{N} \left( 1 + \frac{\alpha}{100} (S_{\text{max}} - 1) \right)
$$

Where the variable $M$ is the size area of the image area, $N$ denotes a gray scale value that is in the range 1-256, $\alpha$ is the clipfactor which denotes adding histograms between 0 and 100 (Desiani et al., 2021).

8) **Median Filter**

Median filter is an image enhancement process using the smoothing method. The median filter can reduce noise by taking a non-linear approach that involves exchanging the center pixel value with the median value of the image matrix (Shah et al., 2022). The median filter operates on the middle element of an $N \times N$ matrix which replaces the value of each pixel with the median value of neighboring pixels that are in a square neighborhood around the pixel being evaluated (Dinç et al., 2015).

9) **One-Hot Encoding**

One-Hot Encoding is a representation method used to convert multicategorical labels into many bicategorical labels. One-Hot vectors consist of binary vectors with a certain length where only one entry has a value of one and the others are 0. Consider a discrete categorical random variable denoted as $x$, which possesses n distinct values, namely $x_1, x_2, ..., x_n$. In the context of one-hot encoding for a specific value $x_i$, it results in a vector $v$. In this vector, all components are set to zero except for the $i$th component, which is assigned the value of 1 (Hancock & Khoshgoftaar, 2020). For example, the segmentation target is a label in the form of a vector with values (0,1,2) indicating A, B, and C, respectively. Vector will be converted successively with one-hot encoding into three vectors, namely (1,0,0) showing A, (0,1,0) showing B, and (0,0,1) showing C (Xing et al., 2023).

### C. **Segmentation**

The segmentation stage is performed by distinguishing objects from the background and eliminating unnecessary parts. The image will be segmented using a CNN with the U-Net architecture to separate features into several parts, namely the nucleus, cytoplasm, and background. U-Net architecture will be augmented with dropout layers and modified to work for multi-class semantic segmentation. In general, the U-Net architecture can be seen in Figure 3.
Figure 3 represents the U-Net architecture for cervical cancer cell semantic segmentation. The U-Net architecture is divided into two parts, the contraction part on the left side and the expansion part on the right side. The contraction part acts as an encoder function, transforming the input image into feature representations at various levels using convolutional blocks and max pooling. The expansion part acts as a decoder function, mapping the acquired features back into pixel density objects through upsampling, merging, and convolution operations. The contraction part consists of 3 blocks, each comprising 2 convolution layers and 1 pooling layer. The expansion part consists of transposed convolution, merging feature maps corresponding to the contraction part, and 2 convolution layers. The ReLU activation function is used in every layer except for the output layer. ReLU is an activation function with better gradient convergence (Rudiansyah et al., 2023). The ReLU activation function is defined in function (5).

$$f(x) = \max(0, x)$$  \hspace{1cm} (5)

ReLU function (5) demonstrates that when the input $x$ is negative, the output remains at 0, and when the input $x$ is positive, the output matches the input (Bai, 2022). In the output layer, the softmax activation function is used to calculate probabilities in multiclass problem (Sharma et al., 2020). Softmax activation function can be defined as equation (6).

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}, \text{where } i = 1, 2, ..., k$$  \hspace{1cm} (6)

The softmax activation function (6) shows the value of the input matrix softmax activation function $i$, $z_i$ is the $i$th input matrix entry and $k$ is the number of classes.

D. Performance Evaluation

Validation results are presented in a confusion matrix. Confusion matrix is a measurement used in evaluating a classification model by classifying the number of predictions of true or false objects. The matrix consists of predictions that will be compared with the original class containing information on actual and predicted values from the classification (Singh et al., 2010). The confusion matrix can be seen in table 1.

<table>
<thead>
<tr>
<th>Class</th>
<th>Prediction Positive</th>
<th>Prediction Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Positive</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Actual Negative</td>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

True Positive (TP) and True Negative (TN) are conditions where the predicted results are the same as the actual situation. False Positive (FP) and False Negative (FN) are conditions where the predicted results are not the same as the actual situation (Shah et al., 2022). From the confusion matrix, accuracy, precision, recall, F1 score, and IoU coefficient values can be determined as model evaluation (Müller et al., 2022).

Mathematically, the evaluation calculation can be seen in equations (8) to (12).
Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)

Precision = \frac{TP}{TP + FP} \quad (9)

Recall = \frac{TP}{TP + FN} \quad (10)

F1 Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (11)

IoU = \frac{TP}{TP + FP + FN} \quad (12)

III. RESULT AND DISCUSSION

1) Data Preprocessing

During data preprocessing phase, it commences by resizing the image to a dimension of 128 x 128 pixels using the resize method. Subsequently, data augmentation is performed on the image, incorporating operations such as horizontal flipping, vertical flipping, sharpening, and gaussian filter. This process yields a total of 4585 images and their corresponding labels, all with dimensions of 128 x 128 pixels. The outcomes of the data augmentation will be further advanced through an image enhancement phase, which includes techniques such as adaptive gamma correction, CLAHE, and median filter.

1) Data Augmentation

Data in the Herlev dataset consists of 917 images along with their corresponding 917 labels. Augmentation processes were performed using four methods such as horizontal flipping, vertical flipping, sharpening, and gaussian blur. Augmentation is applied to each image and label, resulting in 917 new image data and 917 new label data for each method. In total, 3,668 data were generated from the augmentation process. Results from data augmentation process of the 4 methods can be seen in figure 4.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Vertical Flip</th>
<th>Horizontal Flip</th>
<th>Sharpening</th>
<th>Gaussian Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Data augmentation for cervical cell dataset

Figure 4 illustrates that geometrically augmenting the image through the flipping technique yields diverse variations compared to the original image, while the use of the augmentation method involving changes to the pixel values in the image through sharpening kernel and gaussian filter appears as though the image has exclusively enhanced its quality. Data augmentation produces a variety of image variations, thereby increasing the number of images available for processing training data and testing data. After the augmentation process was performed, a total of 4585 data were obtained and utilized in this research.

2) Image Enhancement

Image quality in this research was sequentially enhanced through adaptive gamma correction, CLAHE, and noise reduction using the median filter method. Adaptive gamma correction is used to enhance the contrast of images that are too low. CLAHE and median filter are used with the purpose of improving image quality. An example of the image enhancement process can be seen in figure 5.
Figure 5 shows that the results of image enhancement in the adaptive gamma correction process improve the image contrast so that it looks brighter than before. CLAHE method employed to grayscale the image and enhance the visibility of the nucleus and cytoplasm. However, the CLAHE result introduces a considerable amount of noise, necessitating the application of a median filter to reduce the noise.

3) One-Hot Encoding Labels

Data preprocessing is also performed for labels using the one-hot encoding method. An example of the results of the one-hot encoding process can be seen in Figure 6.

![Figure 6. One-Hot encoding for labels](image)

It can be seen in Figure 4 that the colors used as labels in the semantic segmentation process utilize the outcomes of color alignment to form RGB channels. Segmentation targets in the form of labels which are vectors with values (0,1,2) indicating dark blue, light blue, and red elements respectively will be converted with one-hot encoding into three vectors, namely (1,0,0) indicating dark blue, (0,1,0) indicates light blue, and (0,0,1) indicates red.

2) Image Segmentation

In this research, there are 4,585 images that have been enhanced after data preprocessing. Data will be divided into training data, test data and validation data. Data split consists of 90% training data and 10% test data. Of the 90% of training data, 10% will be taken to become validation data to observe the learning process at the training stage. Image segmentation takes the form of semantic segmentation, specifically involving the labeling of more than two classes. The labels in this research are comprised of three classes, 0 for cytoplasm, 1 for nucleus, and 2 for background. Segmentation was conducted utilizing the U-Net architecture, with a training process spanning 25 epochs and 32 of batch size used. During each epoch, learning was performed on the training data, followed by validation data testing, which yielded accuracy and loss values. The graph of the results of the training process at 25 epochs can be seen in Figure 7.

![Figure 7. Results of training (a)Accuracy, (b)Loss](image)
Figure 7 illustrates the accuracy results from training and testing of validation data always change from each epoch, these results are quite fluctuating, periodically up and down but stable starting from the 20th epoch. The loss graph shows the change in the loss value of the deep learning model during the training process at each epoch. The loss function used is categorical cross-entropy because there are multiclass labels used. Loss graph shown in Figure 7 shows a good decrease in the training process, the loss value and the validation of the loss are not much different. The decline was unstable starting from Epoch 7 and experienced stability at the 20th epoch. Apart from the accuracy and loss values, at the training process, the IoU and RMSE values are also measured. Changes in IOU and RMSE values at each epoch are shown in the figure 8.

![IoU and RMSE Graphs](image)

**Figure 8.** Results of training (a)IoU, (b)RMSE

Based on Figure 8, the IoU value shows an increase from each epoch despite experiencing instability. However, the IoU values for validation during training are very close. At the 25th epoch, the IoU value is close to 0.8, which means that the area formed in each image during the prediction process during training is close to the area on the original label. The RMSE value is also presented in Figure 6 which shows a decrease from the first epoch to the 25th epoch although it is not stable. The RMSE value shows the error obtained from the predicted value and the actual value. The RMSE value in the training process in the last epoch is close to 0.2, which means that the error rate of the model training is quite small. From the results of accuracy, loss, IoU, and RMSE, it was found that the model trained in the training process showed quite good results. Thus, the weights obtained in the training process are used to be tested on the test data. The results of testing on the test data produce performance evaluations in the form of accuracy, recall, precision, and f1-score which are presented in the diagram in figure 9.

![Performance Evaluation Diagram](image)

**Figure 9.** Diagram of performance evaluation results

Based on Figure 9 which contains the results of the performance evaluation from testing the test data using the model formed, it produces good performance, which is above 80%. Evaluation metrics such as accuracy, precision, recall and f1-score values are 89%, 89%, 89% and 88.67% respectively. The results obtained in this study are compared with other studies in table 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** Comparison of our proposed methods with other studies on cervical cancer segmentation
<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-Net &amp; Image Enhancement (Desiani et al., 2021)</td>
<td>77%</td>
<td>-</td>
<td>72%</td>
<td>69%</td>
</tr>
<tr>
<td>U-Net (Zhao et al., 2019)</td>
<td>-</td>
<td>89%</td>
<td>88%</td>
<td>87%</td>
</tr>
<tr>
<td>U-Net &amp; GDLA (G. Li et al., 2022)</td>
<td>-</td>
<td>88.8%</td>
<td>93.6%</td>
<td>-</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>89%</td>
<td>89%</td>
<td>89%</td>
<td>88.67%</td>
</tr>
</tbody>
</table>

Based on a comparison of the results of performance evaluations carried out in other studies on cervical cell image segmentation in Table 2, of the four methods the results produced using this method have precision and recall values that are still smaller than other methods, but using this method produces accuracy and recall values. F1-Score is greater than other methods. This shows quite good results in segmenting the nucleus and cytoplasmic features in cervical cell images. A sample of test results comparing images, labels and prediction results can be seen in Figure 10.

![Figure 10. Comparison of original image with labels and predictions](image)

Based on Figure 10, the prediction results from the segmentation process are compared with the actual image and labels. That figure illustrates the level of similarity of the predicted results with the actual image and label. The prediction results obtained have a fairly high level of similarity with the image and label, meaning that the model can recognize image patterns and features in the image as well. The model formed from the U-Net architecture by adjusting image enhancement using adaptive gamma correction, CLAHE, and median filter can properly segment for the nucleus and cytoplasm at cervical cells dataset.

**IV. CONCLUSION**

Based on the results of research conducted in the process of semantic segmentation in the nucleus and cytoplasm of cervical cancer cells in the Herlev Dataset by adjusting data augmentation techniques and image enhancement, the results are quite good. This is based on an accuracy performance evaluation of 89% which shows a good percentage figure. Apart from accuracy, performance measurements are based on precision, recall, and F1-Score which is around 89% and is a good result.

**BIBLIOGRAPHY**


585


