JOURNAL MEDICAL INFORMATICS TECHNOLOGY
medinftech.org ISSN 2988-7003 (online)

Image Segmentation of Normal Pap Smear Thinprep using U-Net with Mobilenetv2 Encoder

Deviana Sely Wita[1*]

[1]*Universitas Nusa Mandiri Jakarta, Indonesia*

ARTICLE HISTORY

Received: 08 June 23

Final Revision: 10 June 23

Accepted: 20 June 23

Online Publication: 30 June 23

KEYWORDS

Pap Smear, Encoder, U-Net, MobilenetV2, Segmentation

CORRESPONDING AUTHOR

14002609@nusamandiri.ac.id

DOI

10.37034/medinftech.v1i2.6

A B S T R A C T

Pap smear is a technique to detect changes in the cells in the uterine wall. With a Pap smear, a woman can be known to have cervical cancer or not. However, the problem of cancer screening on pap smear images is largely hindered by improper cell staining and overlapping cell images. For accurate Pap smear image segmentation, this study uses the U-Net method which is better for Pap smear image segmentation. This method integrates the MobilenetV2 network and converts ordinary convolution into deep split convolution to improve transmission and feature utilization by the network, and at the same time increase the speed of feature extraction. Then the segmentation results from MobilenetV2 produce accuracy in distinguishing the nucleus, cytoplasm, and background on the pap smear image. The dataset used in this study is a normal analogue image of the Pap smear image obtained from the RepoMedUNM Database. Initial data processing is done by digitizing the image, where analog data from the Pap smear is transformed into a digital image. Based on the results of research that has been carried out, namely segmentation of Pap smear images using U-Net with MobilenetV2 encoder, the accuracy value on differences in nucleus, cytoplasm, and background cells is 98%.

1. Introduction

Cervical cancer is one of the fourth most common and deadly types of cancer among women. In Indonesia, cervical cancer is the second most prevalent type of cancer after breast cancer. The incidence rate of new cervical cancer cases for women in Indonesia is around 36,663, with a death rate of 21,003 [1]. Early detection of cervical cancer can prevent or delay the development of cervical abnormalities into invasive ones [2]. Currently, cervical cytology screening has been widely used, so that cervical cancer and pre-cancerous lesions can be detected and treated early, significantly reducing morbidity and mortality of cervical cancer. Among many methods of cervical cancer screening, Pap smear has been considered as one of the most effective cytology testing methods for early detection of cervical cancer [3].

Pap smear is a technique used to detect changes in cells in the uterine wall. With Pap smear, a woman can be diagnosed with cervical cancer or not. Pap smear is performed by a pathologist manually with the help of a microscope. The time required to identify Pap smear cells is quite long, which is caused by the variety of cells. Image processing needs to be done so that the identification time is faster and the diagnosis errors are smaller [4].

However, the problem of screening cervical cancer in Pap smear images is largely hindered by inaccurate cell staining and overlapping cell images [5]. The analysis involves a lot of human labor to examine hundreds of Pap smear images, requiring a long time [6]. Therefore, computer-based Pap smear image segmentation is needed to help medical practitioners overcome these problems.

Based on the existing U-Net network, a better UNet algorithm is proposed. By replacing the U-Net encoding part with the Mobilenet v2 model, the original ordinary convolution is changed to a depthwise separable convolution. This reduces the network parameters and speeds up the operation. The improved network model proposes an Inverse Residual Block, which reduces the information loss caused by multiple convolutions during sampling, improves the

segmentation results' accuracy, and achieves faster and more efficient feature extraction [7]. This results in more accurate Pap smear segmentation.

1.1. U-Net

The aim of U-Net is to map features from an image (segmentation), this neural network is dedicated to the classification problem for each pixel since the label vector for each pixel is already provided by the radiologist, where each CT scan file can have more than 1 nodule, the label in this case is the nodule area with the Id Finding value that can be seen in the CSV file. The neural network's target is to produce a label vector.

U-Net has an encoder (downsampling) and decoder (upsampling) like Autoencoder. Each encoder layer will reduce the dimension of the input layer, hence it is called downsampling. Each decoder layer will increase the dimension of the input layer, hence it is called upsampling. The thing that differentiates these two types is the existence of a connection layer from the encoder to the decoder (concatenate layer). U-Net usually uses convolutional layers for the encoder and decoder parts rather than dense layers [8].

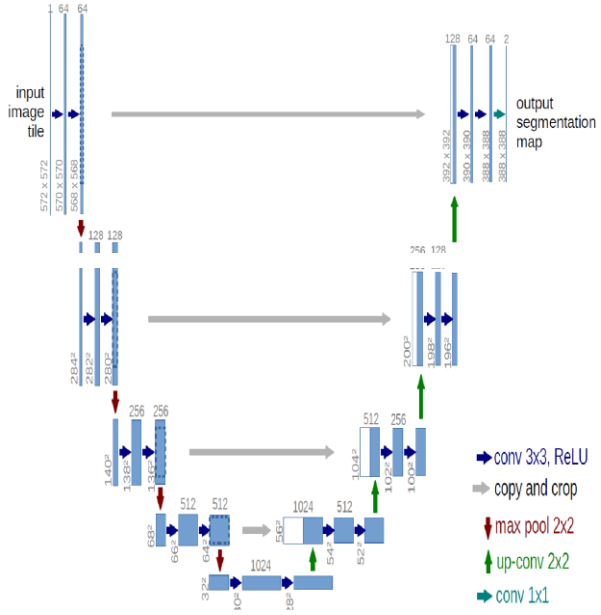

Figure 1. U-Net Architecture [9]

Figure 1 shows the U-Net architecture divided into 3 parts:

a. Contracting/downsampling path

b. Bottleneck

c. Expanding/upsampling path

Contracting or downsampling consists of 4 blocks, and each block is composed of:

a. 3x3 Convolution Layer + activation function

b. 3x3 Convolution Layer + activation function

c. 2x2 Max Pooling

It should be noted that the number of feature maps will double for each pooling, starting with 64 feature maps in the first block, 128 for the second block, and so on. The purpose of contracting is to capture the contextual information of the input image in order to perform segmentation. This rough contextual information will then be transferred to the upsampling path by skip connection. The bottleneck is the part between the upsampling and downsampling paths. The bottleneck is built from only 2 convolutional layers (with batch normalization). Expanding or upsampling aims to enable accurate localization combined with contextual information from the contracting path. It consists of 4 blocks and each block consists of:

a. Deconvolution layer with 2 strides

b. Merging with feature maps cut from the contracting path

c. 3x3 Convolution Layer + activation function (with batch normalization)

d. 3x3 Convolution Layer + activation function (with batch normalization)

1.2. Modification

MobileNet was created by Google and has a special layer called depthwise separable convolution, which is a block consisting of depthwise convolution and pointwise convolution. The purpose of this layer is to reduce computation (to have fewer parameters) so that a smaller model size can be obtained [10]. MobileNet is a small, low-latency, low-power model measured to meet the resource constraints of various use cases. According to research papers, MobileNetv2 improves the performance of advanced mobile models on many tasks and benchmarks as well as across the entire spectrum of different model sizes. MobileNetv2 is a highly effective feature extractor for object detection and segmentation [11].

Using MobileNetv2 in the downsampling part of U-Net, and replacing regular convolution with depthwise separable convolution, results in faster and more accurate pap smear image segmentation.

a. The characteristic of Mobilenetv2 is the Inverted Residual Block. As shown in Figure 2, the number of channels is first increased with 1x1 convolution, and then the spatial 3x3 convolution is converted to Depthwise to achieve the effect of reducing the number of computations. After the dimension is increased, ReLU is added, and then in the dimension in Figure 2.

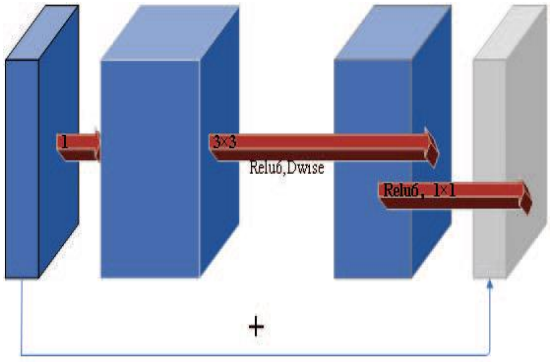

Figure 2. Shows the Inverted Residual Block.

b. The combination of Mobilenet v2 and U-Net network structure is shown in Figure 3. The network structure consists of an encoder, decoder, and skip connection. The encoder part uses the Mobilenetv2 model without any merge operation in the original U-Net network, making the model lighter. The main process is to perform upsampling on f4 with four times length and width compression, then concatenate with f3, then perform upsampling again and f2 on concatenate, then perform upsampling again, and finally use output convolution filter for image class. A total of three upsampling processes are performed in Figure 3.

Mobilenet Encoder			
input			
↓			Conv2d filter = nclasses
Zeropad			BatchNorm
↓			Conv2d
Conv2d s2×2			Zeropad
BatchNorm			↑
ReLU			UpSampling2D
↓			BatchNorm
depthwise_conv_block	→ f1		Conv2d
↓			Zeropad
depthwise_conv_block		f2 →	concatenate
depthwise_conv_block	→ f2		↑
↓			UpSampling2D
depthwise_conv_block			BatchNorm
depthwise_conv_block	→ f3		Conv2d
↓			Zeropad
depthwise_conv_block		f3 →	concatenate
depthwise_conv_block			↑
depthwise_conv_block			UpSampling2D
depthwise_conv_block			BatchNorm
depthwise_conv_block			Conv2d
depthwise_conv_block	→ f4	f4 →	Zeropad
↓			Decoder
depthwise_conv_block			
depthwise_conv_block	→ f5		

Figure 3. U-Net Network Structure Improvement using MobileNetV2 Encoder [7]

2. Research Method

2.1. Research Data

The Pap smear image dataset was obtained from the RepoMedUNM database, consisting of 6139 grouped cell images from Thinprep slides, 2346 and non-Thinprep slides, 3793. Images were obtained using an OLYMPUS CX33RTFS2 optical microscope and an X52-107BN microscope with a Logitech camera.

2.2 Image Segmentation

Pap smear image segmentation uses the U-Net Encoder MobilenetV2 method.

3. Result and Discussion

Testing was performed using a dataset of 70 Pap smear images, divided into 60 training images and 16 validation images. The U-Net Encoder MobilenetV2 algorithm used several hyperparameters and parameters in the Adam Optimizer optimization function. The values of the hyperparameters are as follows: learning rate (Ir) = 0.0001, batch size = 4, verbose = 1, and epochs = 50. The loss chart can be seen in Figure 4 and the accuracy chart in Figure 5.

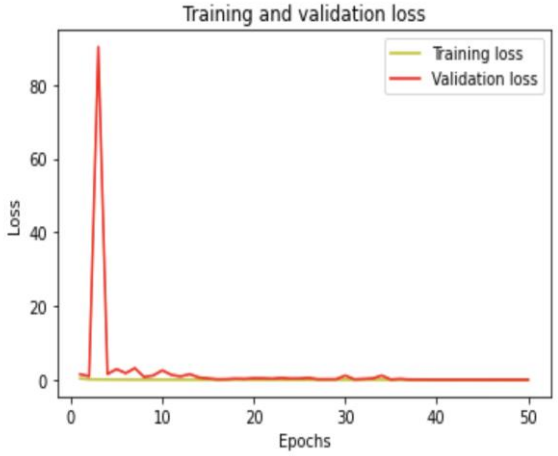

Figure 4. Training and Validation Loss Chart

Figure 4 shows that the training loss has a good value while the validation loss briefly experienced overfitting but stabilized after epoch 10.

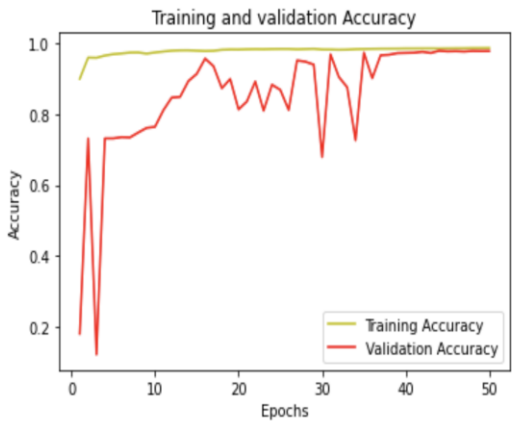

Figure 5. Training and Validation Accuracy Chart

Figure 5 shows that the training accuracy has a good value while the validation accuracy slightly experienced overfitting but after epoch 40, the accuracy value stabilized and improved in Figure 6.

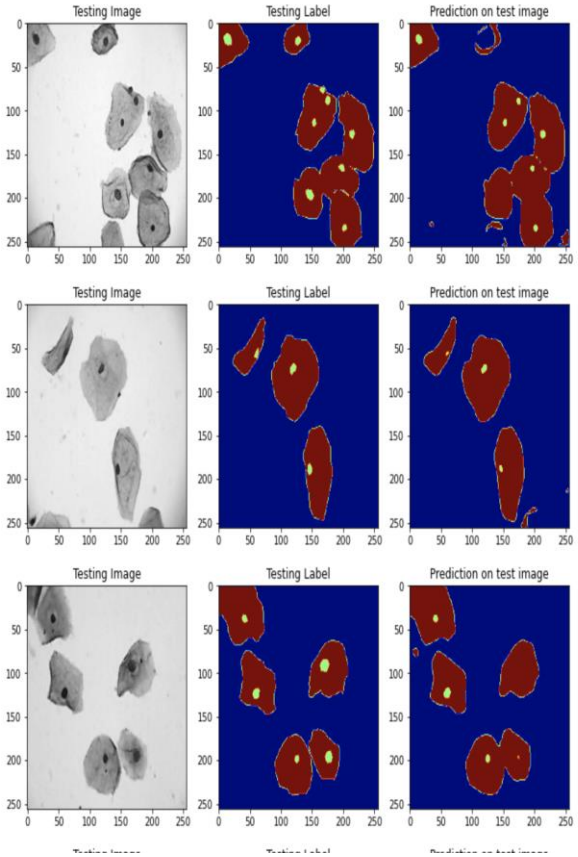

Figure 6. Prediction Result from the U-Net Encoder MobilenetV2 Model

During the process of learning Pap smear image segmentation using the U-Net Encoder MobilenetV2 method, there was no increase in accuracy in epoch 40, so it stopped at that epoch. Figure 6 shows the prediction result from the U-net encoder MobilenetV2 model, where there is a testing image which is the original image, then the testing label is the masked image, and finally the prediction on the test image is the final segmentation result.

Table 1. Accuracy and Loss Results in the Study

Prediction	Value
Epoch	50
Accuracy	98 %
Loss	0.06

From Table 1, the accuracy value at epoch 50 is 98% with a loss value of 0.06.

Testing was performed using a dataset of 70 Pap smear images, divided into 60 training images and 16 validation images. The U-Net Encoder MobilenetV2 algorithm used several hyperparameters and parameters in the Adam Optimizer optimization function. The values of the hyperparameters are as follows: learning rate (Ir) = 0.0001, batch size = 4, verbose = 1, and epochs = 50. The loss chart can be seen in Figure 4 and the accuracy chart in Figure 5.

4. Conclusion

Based on the results of the study, which is the segmentation of Pap smear images using U-Net with the Mobilenetv2 encoder, the accuracy in distinguishing nuclei, cytoplasm, and background in the images is 98%. Therefore, with that accuracy value, it can be concluded that the U-Net encoder Mobilenetv2 method has good accuracy.

References

[1] World Health Organization. (2021). Indonesia Source GLOBOCAN 2020. In International Agency for Research on Cancer." http://gco.iarc.fr/

[2] Ernawati, D. Octaviana, Mantasia, R. Aulia Yusuf, and Sumarmi, "The Effect of Health Education Based on the Health Belief Model about Pap Smear Test on Women in Rural District Indonesia," *MLU*, Mar. 2021, doi: 10.37506/mlu.v21i2.2636.

[3] M. Zhao *et al.*, "SEENS: Nuclei segmentation in Pap smear images with selective edge enhancement," *Future Generation Computer Systems*, vol. 114, pp. 185–194, Jan. 2021, doi: 10.1016/j.future.2020.07.045.

[4] S. Hadianti and D. Riana, "Segmentation and analysis of Pap smear microscopic images using the K-means and J48 algorithms," *J. Teknol. dan Sist. Komput*, vol. 9, no. 2, pp. 113–119, Apr. 2021, doi: 10.14710/jtsiskom.2021.13943.

[5] H. Bandyopadhyay and M. Nasipuri, "Segmentation Of Pap Smear Images For Cervical Cancer Detection," in *2020 IEEE Calcutta Conference (CALCON)*, Kolkata, India, Feb. 2020, pp. 30–33. doi: 10.1109/CALCON49167.2020.9106484.

[6] D. Somasundaram, S. Gnanasaravanan, and N. Madian, "Automatic segmentation of nuclei from pap smear cell images: A step toward cervical cancer screening," *Int J Imaging Syst Technol*, vol. 30, no. 4, pp. 1209–1219, Dec. 2020, doi: 10.1002/ima.22444.

[7] Y. Zhou, M. Chen, M. Zhang, T. Wang, F. Yan, and C. Xie, "Automatic Segmentation of Lung Noudles using improved U-Net NetWork," in *2020 Chinese Automation Congress (CAC)*, Shanghai, China, Nov. 2020, pp. 1609–1613. doi: 10.1109/CAC51589.2020.9326834.

[8] A. Pribadi and A. K. Adisusilo, "Pemanfaatan 3D U-Net untuk Segmentasi 3 Dimensi Gelembung Penyebab Kanker Paru-paru (Nodule) pada Lapisan Citra CT Scan," *INSYST*, vol. 2, no. 2, pp. 74–85, Oct. 2020, doi: 10.52985/insyst.v2i2.159.

[9] I. B. L. M. Suta, M. Sudarma, and I. N. Satya Kumara, "Segmentasi Tumor Otak Berdasarkan Citra Magnetic Resonance Imaging Dengan Menggunakan Metode U-NET," *JTE*, vol. 19, no. 2, p. 151, Dec. 2020, doi: 10.24843/MITE.2020.v19i02.P05.

[10] M. R. R. Allaam and A. T. Wibowo, "Klasifikasi Genus Tanaman Anggrek Menggunakan Metode," p. 37.

[11] N. Nufus *et al.*, "Sistem Pendeteksi Pejalan Kaki Di Lingkungan Terbatas Berbasis SSD MobileNet V2 Dengan Menggunakan Gambar 360° Ternormalisasi," Prosiding Seminar Nasional Sains Teknologi dan Inovasi Indonesia (SENASTINDO), vol. 3, pp. 123–134, Dec. 2021, doi: 10.54706/senastindo.v3.2021.123.