ILKOM Jurnal Ilmiah Vol. 14, No. 3, December 2022, pp. 329-338
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article **Open Access (CC–BY-SA)**

Semantic segmentation of pendet dance images using Multires U-Net Architecture

Hendri Ramdan [a,1,*]; Moh. Arief Soeleman [a,2]; Purwanto [a,3]; Bahtiar Imran [b,4]; Ricardus Anggi Pramunendar [a,5];
[a] *Dian Nuswantoro University, 207 Imam Bonjol Street, Semarang and 50131, Indonesia*
[b] *Mataram Technology of University, Mataram Technology of University Street, Kekalik, Mataram and 83115, Indonesia*
[1] *hendriramdhan8@gmail.com;* [2] *arief22208@gmail.com;* [3] *purwanto@dsn.dinus.ac.id;* [4] *bahtiarimranlombok@gmail.com;*
[5] *ricardus.anggi@dsn.dinus.ac.id;*
* *Corresponding author*

Article history: Received July 11, 2022; Revised October 16, 2022; Accepted November 29, 2022; Available online December 20, 2022

Abstract

As a cultural heritage, traditional dance must be protected and preserved. Pendet dance is a traditional dance from Bali, Indonesia. Dance recognition raises a complex problem for computer vision research because the features representing the dancer must focus on the dancer's entire body. This can be done by performing a segmentation task process. One type of segmentation task in computer vision is the semantic segmentation. Mask R-CNN and U-NET were employed in this task. Since it was first introduced in 2015, semantic segmentation using the U-Net architecture has been widely adopted, developed, and modified. One of the new architectures applied is the MultiRes UNet. This study carries out a semantic segmentation task on the Balinese Pendet dance image using the MultiRes UNet architecture by changing the value of α (alpha) to obtain the best results. This architectural is evaluated by DC score, Jaccard index, and MSE. In this dataset, the alpha value of 1.9 resulted in the best score for DC and the Jaccard index with 98.47% and 99.23% respectively. On the other hand, an alpha value of 1.8 obtained the best score of MSE with 8.20E-04.

Keywords: Semantic Segmentation; U-Net; MultiRes Unet; Deep Learning; Balinese Pendet Dance.

Introduction

As one of the cultural heritages, traditional dance is a crystallization of the nation that must be protected and preserved. It is protected through teaching and exercise among dancers. Unfortunately, this method is considered less effective and is influenced by regional development [1]. Another issue on traditional dance is the threat of being claimed by other countries. In 2009, for example, the Balinese Pendet dance was shown in Enigmatic video documentary produced by KRU studio of Malaysia to promote its tourism. This was adversely affecting the diplomatic relations between Indonesia and Malaysia [2].

Pendet dance is a traditional dance originating from Bali, Indonesia. This dance is a type of sacred dance that is usually performed in temples. But over time, this dance is also used as a dance for entertainment, locally called the balihan mandi dance, and is usually performed to welcome guests. This dance consists of several types of movements. Each movement is a unitary movement of the feet, hands, fingers, body, facial expressions, neck and eyes which makes this dance look more aesthetic [3].

Recent advanced technology, for instance visual recognition and computer vision techniques, can be utilized to preserve the culture particularly traditional dance. However, dances introduce a complex problem for computer vision research as the features that represent a dancer must focus on the dancer's entire body [4]. The segmentation process plays a significant role in introducing dance [4], [5]. A study suggested that thresholding and edge detection methods cannot provide precise segmentation [5]. It is necessary to improve the image quality to improve the quality of the features obtained [6].

In recent years, semantic segmentation has been used in computer vision as a segmentation technique [7]. It is to predict a pixel, whether it belongs to an object or a background. In other words, it is the process of classifying images at the pixel level [8]. Several architectures that can be used in semantic segmentation are Mask R-CNN and U-net [7].

Since it was first introduced in 2015, semantic segmentation using U-Net architecture [9] has been widely used, developed, and modified to perform semantic segmentation tasks on biomedical [8]–[21] and non-biomedical images [7], [22]–[25]. The new architecture including MultiRes Unet architecture [8] generally outperforms the U-Net architecture in performing semantic segmentation tasks.

Based on the description, the dance recognition process has been carried out but does not use semantic segmentation techniques in the segmentation process [4], [5]. In addition, the semantic segmentation task applies the UNet architecture and the modified results do not yet use data dance. Therefore, this study applies Unet's MultiRes architecture for the semantic segmentation task [8] on Balinese Pendet dance images and changes the alpha value to obtain the best results. The architectural measurement used refers to the Dice coefficient (DC) score, the Jaccard index, and the Mean Squared Error (MSE). The contribution of this research is to perform semantic segmentation tasks on Balinese Pendet dance images using Unet's MultiRes architecture based on alpha values.

This study is divided into five chapters. Chapter I describes the study's background, Chapter II presents related research, Chapter III explains the proposed research method, Chapter IV describes the results of the study, and the last chapter V provides conclusions and suggestions for further research.

Method

Several previous studies have proposed different architectures, for instance MultiRes Unet [8]. It aims to obtain the best performance results. However, there is a change in the number of parameters. The alpha value in the MultiRes Block affects these parameters. Therefore, this study is trying to determine the best alpha value to obtain the number of parameters that can describe the best performance. The method proposed by this study is shown by **Figure 1** below.

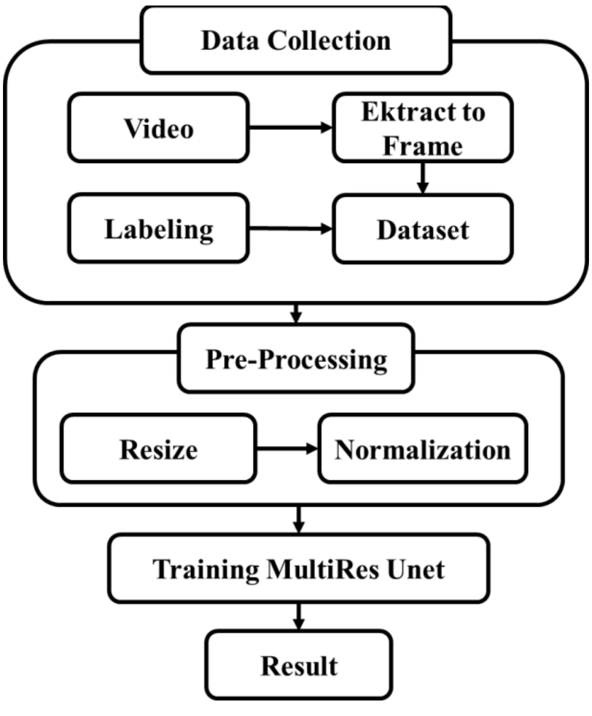
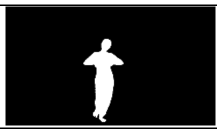


Figure 1. Proposed Method

A. *Data Collection*

The data collected were in the form of video data of 15 types of dance movements consisting of *ngelog*, left *angsel*, *ngeseh*, right *ngangget*, right *sogok*, right *agem*, left *ngangget*, left *sogok*, left *agem*, right *angsel*, right *ngelung*, left *ngelung*, *nyeregseg*, *ngeteb* and throw flowers. Ten dancers demonstrated each type of movement. The collection of video data was then extracted into image data. From the extraction results, five images were taken from each type of dance so that 740 images were obtained. Next, the labeling process was carried out by making ground truth for each image. The total data collected from this process was 1,480 images with a size of 1920 × 1080 pixels. Based on **Table 1** is Examples of Types of Dance Movements.

Table 1. Examples of Types of Dance Movements

Original Image	Ground Truth	Movement Type
		*Ngelog*
		Left *Angsel*
		*Ngeseh*
		Right *Ngangget*
		Right *Sogok*

B. Pre-Processing

At this stage, the images were resized into 256 × 192 pixels. After that, the normalization process of the dataset was carried out. Normalization aims to change the pixel value by dividing each pixel by 255 to produce a value between 0 and 1 [26].

C. MultiRes UNet

MultiRes UNet is one of the new architectures, an improvement of the UNet architecture. This architecture replaces two layers of encoder-decoder convolution with multires blocks for each layer. Each multires block consists of three layers of 3x3 convolutions, which are layered and combined to describe the spatial feature scale. In addition, a 1x1 convolution is added as a residual connection from input to output to add spatial information [8], [27]. The multires Block is shown in **Figure 2** below.

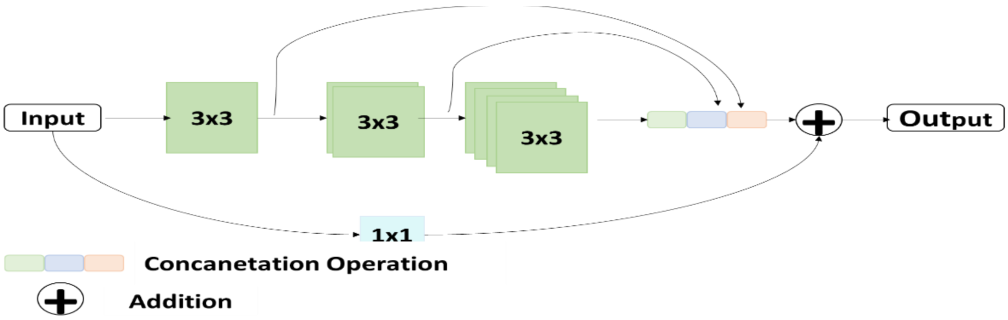


Figure 2. Multi-Res Block

Parameters are assigned in each Block to control the number of convolution filters inside the Block. The number of parameters is calculated using the following **Equation 1** [8]:

$$W = \alpha \times U \tag{1}$$

Here, W is the number of parameters, α or alpha is the scaler coefficient, and U is the number of filters. In addition, this architecture replaces shortcut paths with res paths to reduce feature representation differences between encoder and decoder layers consisting of 3x3 and 1x1 convolutions [8], [27]. This res path is shown in **Figure 3**.

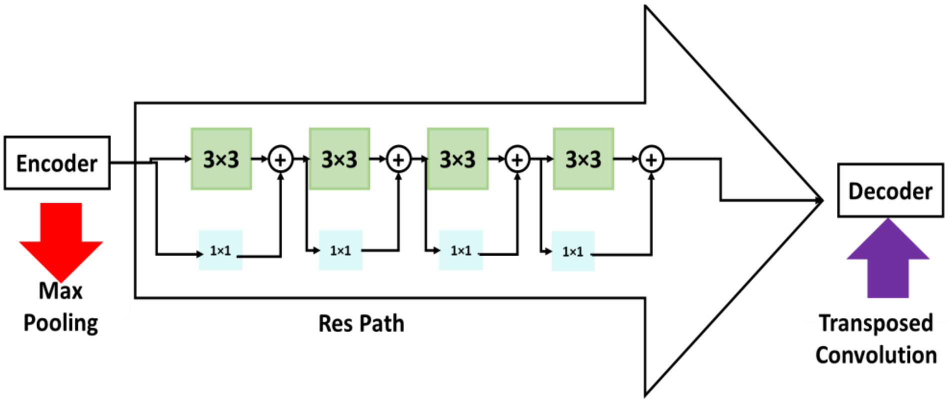


Figure 3. Res Path

In this architecture, the Rectified Linear Unit (ReLU) activation function was used to activate all convolution layers in all networks except the output layer. ReLU is determined using the following **Equation 2** [28]:

$$ReLU = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

Here, x is a feature in the input image. Meanwhile, the loss function binary cross-entropy (BCE) is used to train the model and is defined as follows **Equation 3** [29]:

$$BCE(X, Y, \acute{Y}) = \sum_{px \in X} -(Ypx \, \log(\acute{Y}px) + (1 - Ypx) \log(1 - \acute{Y}px)) \quad (3)$$

X is the input image, Y is the ground truth image, Ý is the segmentation result, px is the pixel, Ypx is the ground truth value, and Ýpx is the model prediction. The MultiRes UNet architecture is illustrated by **Figure 4** [8].

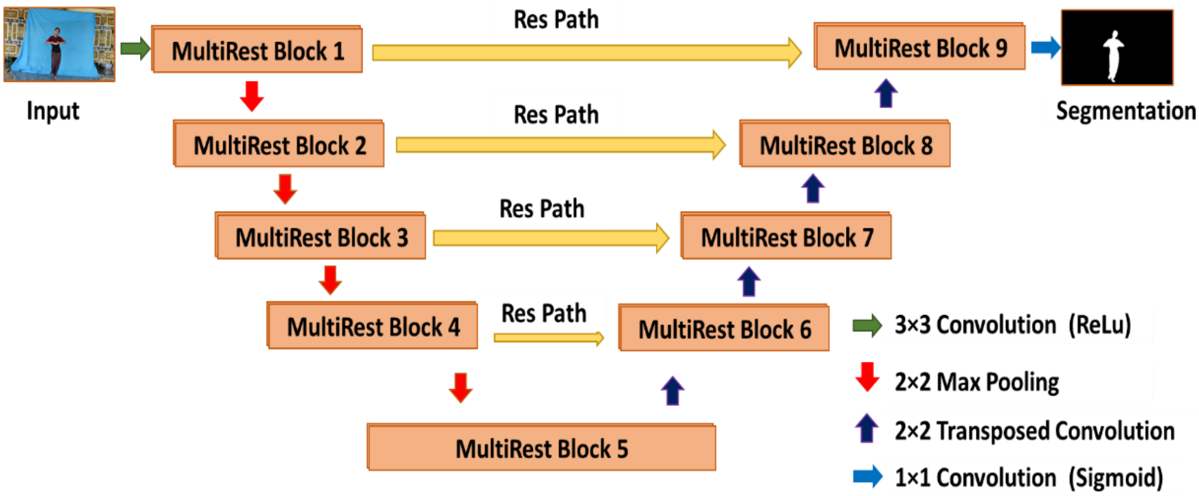


Figure 4. MultiRes UNet Architecture

D. Evaluation

Three evaluation methods were used to evaluate the MultiRes UNet architecture: Dice Coefficient (DC), Jaccard Index, and Mean Square Error (MSE).

1. Dice Coefficient

The dice coefficient is also known as the F1 score or dice similarity coefficient (DSC). It is usually used to evaluate segmentation tasks and semantic segmentation based on the level of similarity. To calculate the dice coefficient, the following **Equation 4** is applied [15]:

$$\mathrm{DC} = \frac{2\ |A \cap B|}{|A| + |\ B|} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

A is the mask image, B is the result of segmentation, TP is genuinely positive, FP is a false positive, and FN is a false negative.

2. Jaccard Index

The Jaccard index has been widely used in semantic segmentation tasks. It is used to compare the mask image with the segmentation results to find the level of similarity between two pictures. The value of the Jaccard index is determined using **Equation 5** as follows [8].:

$$\mathrm{JaccrdIndex} = \frac{A\ Intersection\ B}{A\ Union\ B} = \frac{A \cap B}{A \cup B} \quad (5)$$

3. Mean Squared Error (MSE)

The mean squared error is used to measure the error rate in the segmented image. To calculate the MSE score, the following **Equation 6** is used [30].:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=0}^{n} (A_i - B_i)^2 \quad (6)$$

n is the product of the height and width of the image, while i is the pixel value of the image.

Results and Discussion

A. Tools

The experiment was run using the Python programming language and Google Colab. The specification of the device used is the AMD Ryzen 3 5000 series processor with 8GB RAM and AMD Radeon Graphics.

B. Dataset and Pre-Processing

Table I shows the pendet dance dataset used with a total of 1480. All images in this dataset were resized into 256 x 192 pixels and then normalized. The results of resizing and normalization can be seen in **Table 2** and **Table 3**.

Table 2. Example Of a Resized Image

Original Image	Image Resize	Movement Type
		*Ngelog*
		Left *Angsel*
		*Ngeseh*
		Right *Ngangget*
		Right *Sogok*

Table 3. Image Normalization Example

Original Pixel Value	Normalized Pixel Value
[[[0 0 18] [19 37 57] [53 96 122] ... [210 235 236] [184 209 208] [156 180 181]] [[76 125 147] [56 113 139] [14 80 107] ... [125 152 171] [169 196 207] [192 221 225]]	[[[0. 0. 0.07058824] [0.0745098 0.14509804 0.22352941] [0.20784314 0.37647059 0.47843137] ... [0.82352941 0.92156863 0.9254902] [0.72156863 0.81960784 0.81568627] [0.61176471 0.70588235 0.70980392]] [[0.29803922 0.49019608 0.57647059] [0.21960784 0.44313725 0.54509804] [0.05490196 0.31372549 0.41960784] ... [0.49019608 0.59607843 0.67058824] [0.6627451 0.76862745 0.81176471] [0.75294118 0.86666667 0.88235294]] [[0.08235294 0.33333333 0.47058824] [0.16470588 0.36078431 0.48235294] [0.16862745 0.32941176 0.43921569] ... [0.04705882 0.19607843 0.40784314] [0.07058824 0.23529412 0.41568627] [0.10980392 0.2745098 0.44705882]]

C. Results

During model training, the alpha values in each MultiRes Block were changed. This alpha value affects the number of parameters on the architecture used. The number of parameters for each alpha is shown in **Table 4**.

Table 4. Number of Parameters of Each Alpha

Alpha Value	Number of Parameters
1.1	4.061.660
1.2	4.564.129
1.3	5.083.145
1.4	5.632.863
1.5	6.236.317
1.6	6.833.213
1.7	7.457.910
1.8	8.109.133
1.9	8.801.226
2.0	9.527.752

Table 4 shows that the higher the alpha values, the more the number of parameters. The highest number of parameters is 9,527,752 with an alpha value of 2.0, while the lowest number of parameters is 4,061,660 with an alpha value of 1.1. The training process for each alpha value was performed ten times then the results were averaged to find the best score of DC, Jaccard index, and MSE, as shown in **Table 5** below.

Tabel 5. Mean Score Of DC, Jaccard Index, and MSE

Alpha Value	DC	Jaccard Index	MSE
1.1	98.05	99.00	1.19E-03
1.2	98.29	99.12	1.00E-03
1.3	98.29	99.14	1.04E-03
1.4	98.42	99.20	1.02E-03
1.5	98.36	99.17	9.59E-04
1.6	98.25	99.12	9.65E-04
1.7	98.24	99.11	9.17E-04
1.8	98.46	99.22	**8.20E-04**
1.9	**98.47**	**99.23**	8.68E-04
2.0	98.41	99.20	1.04E-03

Figure 5 consists of the X-axis and Y-axis, where the X-axis is the alpha value while the Y-axis is the DC score. **Figure 5** shows that the highest DC score is 98.47 which was obtained by an alpha value and the number of parameters of 1.9 and 8,801,226 respectively. In contrast, the lowest score is 98.05, obtained by an alpha value of 1.1 and a total of 4,061,660 parameters, the lowest parameter number. Furthermore, the Jaccard index score can be seen in **Figure 4**.

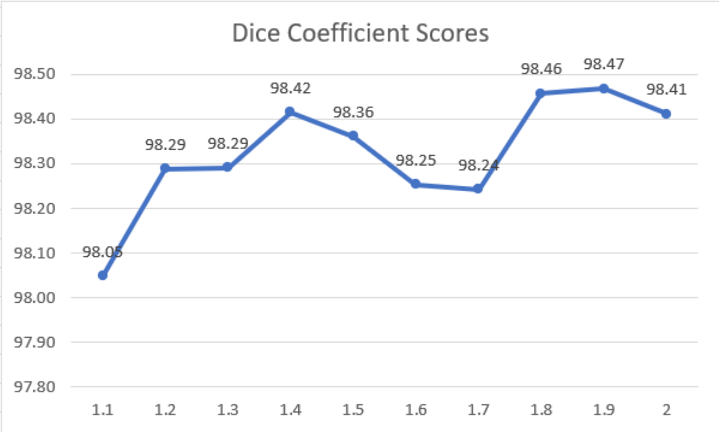


Figure 5. Dice Coefficient Score Chart

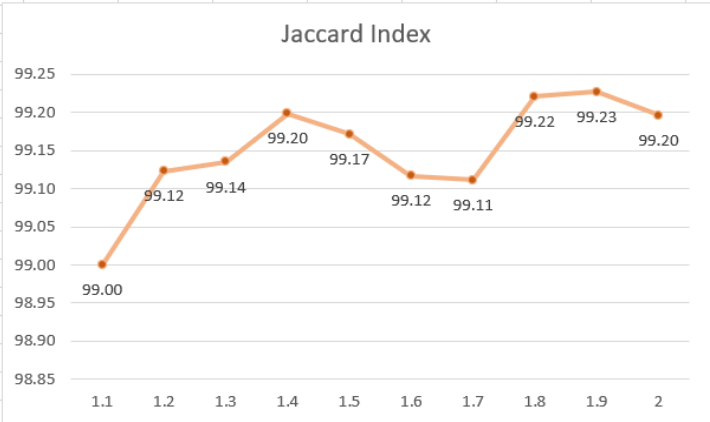


Figure 6. Jaccard Index Score Chart

Figure 6 also consists of the X axis and Y axis, where the X axis is the alpha value, while the Y axis is the jaccard index score. Similar to the DC score, the highest score of the jaccard index is 99.23, obtained by an alpha value of 1.9 and the number of parameters of 8,801,226. Furthermore, the lowest score on the Jaccard index was obtained by an alpha value of 1.1 with a score of 99.00 with a total of 4,061,660 parameters which is the lowest number of parameters. While the MSE score can be seen in **Figure 7**.

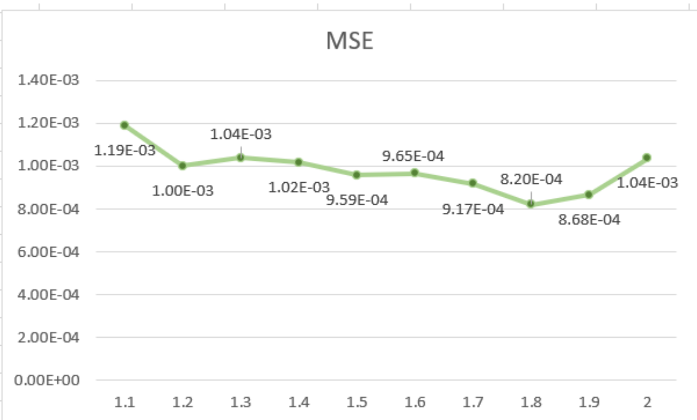
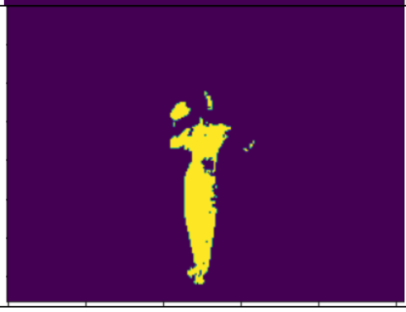


Figure 7. Mean Squared Error Score Chart

Likewise, **Figure 7** also consists of the X-axis and Y-axis, where the X-axis is the alpha value while the Y-axis is the MSE score. In the MSE score, the alpha value of 1.1 gets the highest error value with a score of 1.19E-03. While the MSE score that gets the lowest error value is an alpha value of 1.8 with a score of 8.20E-04 and the number of parameters is 8109.133, where this score is not directly proportional to the DC score and the jaccard index which gets the highest score at an alpha value of 1.9. Because if seen from the last three error rate scores from the experiments conducted, it shows that after the MSE score gets the lowest error value and after that the number of parameters increases then will cause the error value to increase as well. This affects the segmentation results obtained, as seen in **Table 6**.

Likewise, **Figure 7** also consists of the X axis and Y axis, where the X axis is the alpha value while the Y axis is the MSE score. The highest MSE score is 1.19E-03, obtained at an alpha value of 1.1 while the lowest score is 8.20E-04, obtained at an alpha value of 1.8 and the number of parameters of 8109.133. This score is not directly proportional to the DC and the jaccard index score which show the highest score at an alpha value of 1.9. From the last three error rate scores in the experiments conducted, if the number of parameters increases after the lowest MSE value is obtained, the error value will also increase. This affects the segmentation results obtained as shown in **Table 6**.

Table 6. Segmentation Results

Alpha Value	Ground Truth	Segmentation Results
1.8		
1.9		

Conclusion

Based on the results of the experiments conducted, it can be concluded that for both DC and jaccard index, the best score was obtained by an alpha value of 1.9 and total parameter of 8,801,226 with a score of 98.47 and 99.23 respectively. Meanwhile, the best MSE score was obtained by an alpha value of 1.8 and total parameter of 8,109,133 with a score of 8.20E-04; the lowest score was obtained by an alpha value of 1.1 with the lowest number of parameters.

This study proposes that during the model training process, each alpha value should be repeated ten times. In addition, since the alpha value being tested is still in a range from 1.1 to 2.0, the future researches are suggested to use k-fold cross-validation during the training process and use optimization techniques to try more alpha values in order to obtain the best value. Moreover, the lowest error rate obtained must follow the DC score and the highest jaccard index.

References

[1] C. Cui, J. Li, D. Du, H. Wang, P. Tu, and T. Cao, "The method of dance movement segmentation and labanotation generation based on rhythm," *IEEE Access*, vol. 9, pp. 31213–31224, 2021, doi: 10.1109/ACCESS.2021.3060103.

[2] L. Y. MENG and Y. KARULUS, "the Indonesia-Malaysia cultural heritage disputes: a case study of the

pendet dance and rasa sayange folk song," *MANU J. Pus. Penataran Ilmu dan Bhs.*, no. July, 2019, doi: 10.51200/manu.v0i0.1880.

[3] G. A. M. Puspawati and L. De Liska, "Nilai-Nilai Pendidikan Karakter dalam ragam Gerak tari Pendet," *Stilistika J. Pendidik. Bhs. dan Seni*, vol. 7, no. 2, pp. 274–291, 2019, doi: 10.5281/zenodo.3900648.

[4] K. V. V. Kumar, P. V. V. Kishore, and D. Anil Kumar, "Indian classical dance classification with adaboost multiclass classifier on multifeature fusion," *Math. Probl. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/6204742.

[5] K. V. V. Kumar and P. V. V. Kishore, "Indian classical dance mudra classification using HOG features and SVM Classifier," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 5, pp. 2537–2546, 2017, doi: 10.11591/ijece.v7i5.pp2537-2546.

[6] R. A. Pramunendar, D. P. Prabowo, D. Pergiwati, Y. Sari, P. N. Andono, and M. A. Soeleman, "New workflow for marine fish classification based on combination features and CLAHE enhancement technique," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, pp. 293–304, 2020, doi: 10.22266/IJIES2020.0831.26.

[7] F. Sun, V. Ajith Kumar, G. Yang, A. Zhang, and Y. Zhang, "Circle-u-net: An efficient architecture for semantic segmentation," *Algorithms*, vol. 14, no. 6, 2021, doi: 10.3390/a14060159.

[8] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020, doi: 10.1016/j.neunet.2019.08.025.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[10] Q. Zuo, S. Chen, and Z. Wang, "R2AU-Net: attention recurrent residual Convolutional Neural Network for multimodal medical image segmentation," *Secur. Commun. Networks*, vol. 2021, 2021, doi: 10.1155/2021/6625688.

[11] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari1, "Recurrent residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," 2018, doi: 10.1007/978-3-030-47436-2_16.

[12] V. Ashkani Chenarlogh *et al.*, "Clinical target segmentation using a novel deep neural network: double attention Res-U-Net," *Sci. Rep.*, vol. 12, no. 1, pp. 1–17, 2022, doi: 10.1038/s41598-022-10429-z.

[13] H. Cao *et al.*, "Swin-Unet: Unet-like pure transformer for Medical Image Segmentation," pp. 1–14, 2021, [Online]. Available: http://arxiv.org/abs/2105.05537

[14] D. Maji, P. Sigedar, and M. Singh, "Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors," *Biomed. Signal Process. Control*, vol. 71, no. PA, p. 103077, 2022, doi: 10.1016/j.bspc.2021.103077.

[15] K. B. Soulami, N. Kaabouch, M. N. Saidi, and A. Tamtaoui, "Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation," *Biomed. Signal Process. Control*, vol. 66, no. November 2020, p. 102481, 2021, doi: 10.1016/j.bspc.2021.102481.

[16] Z. Wang, Y. Zou, and P. X. Liu, "Hybrid dilation and attention residual U-Net for medical image segmentation," *Comput. Biol. Med.*, vol. 134, no. April, p. 104449, 2021, doi: 10.1016/j.compbiomed.2021.104449.

[17] K. Sanjar, O. Bekhzod, J. Kim, J. Kim, A. Paul, and J. Kim, "Improved U-net: Fully convolutional network model for skin-lesion segmentation," *Appl. Sci.*, vol. 10, no. 10, pp. 1–14, 2020, doi: 10.3390/app10103658.

[18] N. S. Punn and S. Agarwal, "Inception U-Net Architecture for Semantic Segmentation to Identify Nuclei in Microscopy Cell Images," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 16, no. 1, 2020, doi: 10.1145/3376922.

[19] Y. Cai and Y. Wang, "MA-Unet: an improved version of Unet based on multi-scale and attention mechanism for medical image segmentation," p. 16, 2022, doi: 10.1117/12.2628519.

[20] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-Net: Multi-scale densely connected U-Net for biomedical image segmentation," 2018, [Online]. Available: http://arxiv.org/abs/1812.00352

[21] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-Unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, no. c, pp. 44247–44257, 2019, doi: 10.1109/ACCESS.2019.2908991.

[22] S. Piao and J. Liu, "Accuracy Improvement of UNet Based on Dilated Convolution," in *Journal of Physics: Conference Series*, 2019, vol. 1345, no. 5. doi: 10.1088/1742-6596/1345/5/052066.

[23] A. Abdollahi, B. Pradhan, and A. M. Alamri, "An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images," *Geocarto Int.*, vol. 0, no. 0, p. 000, 2020, doi: 10.1080/10106049.2020.1856199.

[24] M. Lei, Z. Rao, H. Wang, Y. Chen, L. Zou, and H. Yu, "Maceral groups analysis of coal based on semantic segmentation of photomicrographs via the improved U-net," *Fuel*, vol. 294, no. June 2020, p. 120475, 2021, doi: 10.1016/j.fuel.2021.120475.

[25] Z. Wang *et al.*, "Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery," *Remote Sens.*, vol. 14, no. 1, 2022, doi: 10.3390/rs14010045.

[26] B. Imran, "Classification Of lombok songket cloth image using Convolution Neural Network method (CNN)," no. 85, pp. 149–156, 2020, doi: 10.33480/pilar.v17i2.2705.

[27] K. Armanious, V. Kumar, S. Abdulatif, T. Hepp, S. Gatidis, and B. Yang, "IpA-MedGAN: Inpainting of arbitrary regions in medical imaging," in *Proceedings - International Conference on Image Processing, ICIP*, 2020, vol. 2020-Octob, pp. 3005–3009. doi: 10.1109/ICIP40778.2020.9191207.

[28] C. Yan, X. Fan, J. Fan, and N. Wang, "Improved U-Net remote sensing classification algorithm based on multi-feature fusion perception," pp. 1–18, 2022.

[29] A. Abdollahi and B. Pradhan, "Integrating semantic edges and segmentation information for building extraction from aerial images using UNet," *Mach. Learn. with Appl.*, vol. 6, no. October, p. 100194, 2021, doi: 10.1016/j.mlwa.2021.100194.

[30] Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, "Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks based U-Net," *NAECON 2018 - IEEE Natl. Aerosp. Electron. Conf.*, pp. 228–233, 2018.