Segmentation of Microaneurysms for Early Detection of Diabetic Retinopathy Using MResUNet

Dinial Utami Nurul Qomariah[1]   Handayani Tjandrasa[1]*   Chastine Fatichah[1]

[1]Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
* Corresponding author's Email: handatj@its.ac.id.

Abstract: Diabetic retinopathy is a disease caused by complications of chronic diabetes and can lead to blindness. Therefore, automatic detection of microaneurysms on retinal fundus images as an early warning system is very necessary and challenging since the size of the microaneurysm is very small compared to the anatomical structures of the retina such as blood vessels and other elements, and hemorrhages as other red lesions. In this study, we propose a novel deep learning network that modifies UNet using residual units with modified identity mapping (MResUNet) to perform microaneurysm segmentation. The purpose of identity mapping modification using convolutional layers and batch normalization is to enrich features and add Residuals in UNet to overcome feature degradation as the network becomes deeper. The mean weighted loss function is used in training to solve the problem of pixel imbalance between the microaneurysms and the background. The proposed architecture is evaluated using the IDRID and DiaretDB1 datasets. The experimental results show that the proposed architecture (MResUNet) achieves higher sensitivity values of 61.96% and 85.87% on the IDRID and DiaretDB1 datasets, respectively, compared to AutoEncoder, FCN16, FCN8, and UNet.

Keywords: Deep learning, Segmentation, Microaneurysms, Diabetic retinopathy, Residual network, UNet.

1. Introduction

Chronic diabetes can lead to diabetic retinopathy of the retina [1]. Diabetic retinopathy can cause blindness in productive age [2]. The severity of diabetic retinopathy involves non-proliferative [3-5] and proliferative diabetic retinopathy [6, 7]. Non-proliferative diabetic retinopathy is divided into three stages which are mild, moderate, and severe non-proliferative [8]. Mild non-proliferative diabetic retinopathy is characterized by abnormalities in the blood vessels, indicated by the appearance of microaneurysms. Microaneurysms can rupture causing hemorrhages. Microaneurysms and dot hemorrhages have the same small shapes with the appearance of blackish-red dots called red small dots [8-10]. Both lesions are early signs of diabetic retinopathy [11]. Therefore automatic detection of microaneurysms is very important for early detection of diabetic retinopathy [12].

Although there have been many researches to detect lesions such as microaneurysms, hard exudates, soft exudates, and hemorrhages [13, 14], lesion segmentation, especially for microaneurysms, is still challenging research this time, because the size is very small compared to the background, only about one percent of the whole image [8] with contrast values that resemble the background, blood vessels, and hemorrhages. Recent researches employed deep learning methods for medical image segmentation, such as CNN transfer learning, because they can produce high accuracy. Several studies proposed using deep learning for microaneurysm detection [15-18]. Most of the researches used image patching and deep network architectures to detect microaneurysms [12, 15, 16], and proposed multi-scale patch-wise refinement to segment microaneurysms with triplet loss [18]. Sarhan et al. [18] proposed two different architecture stages using hypothesis generation network and patch wise

refinement network. Using a different network causes time consuming training times.

Recent researches on segmentation using deep learning-based semantic segmentation have been carried out, such as fully convolutional network (FCN) [19], AutoEncoder [20], and UNet [21]. UNet is widely used in medical images because of its ability for segmentation and detection [21]. UNet uses low-level and high-level architectures to collect information on objects, but deeper networks can suffer from feature degradation. Residual network [22] overcomes feature degradation with the skip connection, identity mapping for feature propagation.

Therefore, in this paper we propose Modified UNet-Residual to detect microaneurysm by applying a modified identity mapping and Residual to UNet architecture, to obtain microaneurysm segmentation. Identity mapping modification uses convolutional layers and batch normalization to enrich features, so that the features that are lost when passing through the plain Unet are replaced with features from identity mapping modifications. We use an adaptive weighted cross entropy loss function on the MResUnet to solve the problem of pixel imbalance between microaneurysms and the background. The weight values are calculated based on the mean value of the pixel frequencies of the training image. The weighted loss function can improve the network architecture to detect object classes that are minority classes and can also be applied to multiclass segmentation. Our proposed MResUNet is evaluated using the IDRiD dataset [23] which has pixel imbalance between microaneurysms and background. In addition, MResUNet is also tested with DiaretDB1 images [10] to prove that the proposed model can be applied to retinal fundus images of another dataset. This research compares with state-of-the-art deep learning methods to measure the performance of the proposed network architecture.

The rest of this paper is organized as follows. Section 2 gives the related work on microaneurysm detection. Section 3 describes the proposed metodology for microaneurysm segmentation using MResUnet. Section 4 discusses the experiments performed and analyzes the results. Section 5 provides the conclusion and the future work of the research.

2. Related works

Recent studies on segmentation using deep learning based semantic segmentation have been done, including FCN, AutoEncoder, and UNet. Using FCN [19] the input image passes in the convolution process and produces a direct output, this process accelerates the computation process. However, FCN lacks detail to generate the output because the process produces the output immediately. The problems can be overcome by AutoEncoder [20], such as VGG auto encoder. VGG AutoEncoder has a network that performs encoder and decoder that produced output more detailed compared to FCN. However, VGG AutoEncoder cannot determine detail between edges and objects, especially in medical images. To overcome this problem, the modification of the auto encoder is UNet. UNet is an architecture that is widely used in medical images [21], because of its ability to detect the edges of objects. The difference between UNet and AutoEncoder is the skip connection on UNet which forwards information on the encoder block to the decoder block to restore spatial information on the output network.

Several studies have been conducted to detect lesions in fundus images, such as using deep learning. Deep learning in recent years is a popular method because of its outstanding performance, especially on medical images [18, 24-26]. Hong et al. designed a CNN architecture consisting of 10 layers to detect exudates, hemorrhages, and microaneurysms [25]. Each pixel is grouped on a probabilistic basis, this results in high computation time. Guo et al. modified the architecture of the VGG auto encoder to perform segmentation of hard exudates, soft exudates, hemorrhages, and microaneurysms. Segmentation of 4 lesions simultaneously makes the detection results of the four lesions less successful because the comparison of each class in each lesion is different [26]. Sarhan et al. proposed a 2-stage network to determine the areas of the microaneurysms and perform patch extraction on the microaneurysms [18]. Since microaneurysms are lesions that early signs of diabetic retinopathy, this study was proposed to detect microaneurysms lesions. Microaneurysms became difficult to detect because of the imbalanced pixel ratio between microaneurysms and background. Shown in Table 1, microaneurysms pixels on the IDRiD dataset are far less than the background pixels [26]. Therefore, we need a weighted loss function, such as a cross entropy loss function. Weighted cross entropy allows the network to detect microaneurysms properly. The weighted cross entropy loss function shows performance improvement in fundus lesion segmentation and is able to overcome the imbalanced class problem in microaneurysms.

Therefore, in this study, we propose MResUnet to perform microaneurysm segmentation. Our proposed network architecture combines the advantages of UNet which facilitate end-to-end training and Residuals that facilitate the training process as well
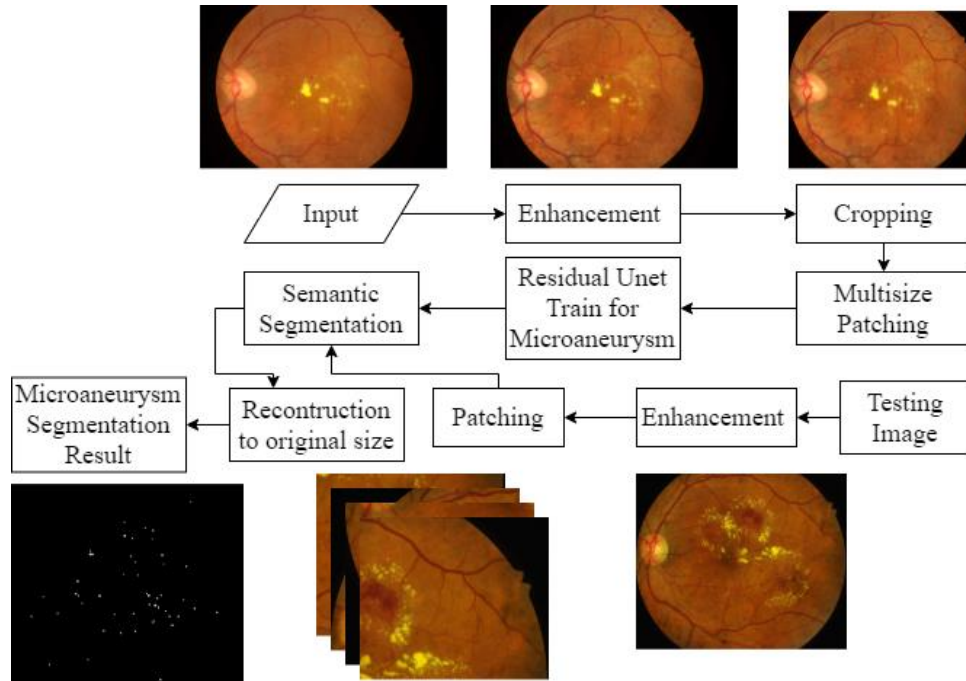
Figure. 1 The research diagram shown using the IDRiD images

Table 1. Statistical information of the percentage microaneurysm pixels for the IDRiD dataset.

Dataset	Microaneurysms	Total Images
IDRiD	0.10%	81

as a modified identity mapping which propagates the information that allows us to have a network design that has few parameters but has good performance. We use the adaptive weighted cross entropy loss function for the MResUNet network to overcome the problem of imbalanced pixel ratio between the object and the background. The weight values are calculated based on the mean value of the pixel frequencies of the training images. The weighted loss function can improve the network architecture to detect object classes that are minority classes.

3. Material and methodology

This research performs segmentation between microaneurysms and background using the IDRiD dataset [23]. This research has 5 stages, they are enhancement, cropping, patching, training using MResUNet, and testing. The research diagram is shown in Fig. 1.

3.1 Dataset

We used IDRiD and DiaretDB1 datasets. The IDRiD dataset, is a new publicly available database, was obtained from a clinic in the Nanded area of Maharashtra, India [23]. For annotation pixels, Idrid

has 81 images that have signs of diabetic retinopathy. Each image is provided with microaneurysms ground truth. Idrid dataset was taken using a VX-10 alpha kowa digital fundus camera with a 50-degree field of view and centered on the macula area. Each image is saved in .jpg format and has a size of 4288×2848 with a color image.

The DiaretDB [10] is a public database for benchmarking diabetic retinopathy detection. The database was obtained from Kuopio University Hospital. The dataset has 89 images with 84 signs of diabetic retinopathy and 5 no signs of diabetic retinopathy. An image was taken using a 50 degree field of view digital fundus camera. The images were selected by a medical expert with the ground truths of the red small dots.

3.2 Image enhancement

Enhancement methods used in this study were cumulative histogram equalization and contrast limited adaptive histogram equalization (CLAHE). The purpose of this stage is to increase the contrast in the image so that the microaneurysms become more visible that can be detected properly.

Our training image enhancement uses cumulative histogram equalization to improve image quality. The enhancement process is applied to each channel, then returned to an RGB image.

For testing images, we apply CLAHE to each channel to increase the contrast value of the test image, CLAHE can help the microaneurysm area

become more visible so it can produce properly the segmentation of the microaneurysm area in the test image. After enhancing in each channel, the result is returned to an RGB image.

3.2.1. Cumulative histogram equalization

Cumulative histogram equalization is an image enhancement that can improve the image contrast by modifying the histogram intensity distribution. The grayscale image is denoted by g and n_i is the number of occurrences of the gray value i [27], then the probability of occurrence of the pixel value i in the image is calculated as in Eq. (1).

$$p_g(i) = p(g = i) = \frac{n_i}{n},$$
$$0 \le i \le L - 1 \qquad (1)$$

where $L - 1$ is the maximum level in a grayscale image. The total number of pixels in the image is n, and $p_g(i)$ becomes the image histogram value for the pixel value i, normalized in the range of [0,1].

The cumulative distribution function $CDF_g(i)$ corresponding to $p_g(j)$ is calculated in Eq. (2).

$$CDF_g(i) = \sum_{j=0}^{i} p_g(j). \qquad (2)$$

Below are the steps for implementing the cumulative histogram equalization [28].
1. Calculate the image histogram.
2. Calculate the cumulative distribution function.
3. Calculate the new gray values based on the general formula for the histogram equalization.
4. Assign the new values for each grayscale value in the image.

3.2.2. Contrast limited adaptive histogram equalization (CLAHE)

CLAHE is one of the equalization methods widely used in the image enhancement. CLAHE can improve image contrast between microaneurysms and background [29]. CLAHE divides the image into tiles and performs histogram equalization for each pixel value in the image [30].

$$n_{avg} = \frac{n_{cr-x} \times n_{cr-y}}{n_{gray}}, \qquad (3)$$

where n_{avg} is the average number of pixels, n_{gray} represent the number of gray levels in the contextual

region, n_{cr-x} denotes the number of pixels in the x direction in the contextual region, and n_{cr-y} is the number of pixels in the y direction in the contextual region.

After that, calculate the actual clip limit value in Eq. (4).

$$n_c = n_{clip} \times n_{avg} \qquad (4)$$

where n_c represents the clip limit and n_{clip} denotes the maximum average pixel value of each gray degree in the sub image.

Following are the steps for implementing CLAHE [27].
1. Partition the image into sub-images with size m × n.
2. Calculate the histogram for each sub-image.
3. Calculates clip histogram for each sub-image.
4. Image enhancement is carried out by interpolating gray level mapping.

3.3 Cropping

Before the patching stage, we cropped the image that had been enhanced. Cropping was done by taking the retina area and reducing the background because the background area is dominant to the retina area.

3.4 Patching

Deep learning is a learning method that requires many training images to get the appropriate results. Therefore, due to the limited number of images, we patch the training image. In addition, because the microaneurysm lesion is very small around the blood vessels, patching is needed to maintain the quality of the image. The idea of using patching can increase the amount of training data. For patching the training images, we choose the parts of images that contain microaneurysms. To increase the detection accuracy of the microaneurysms, patching with one size is still insufficient to represent the areas of the microaneurysms. Thus, we propose multi-size patching to increase the accuracy of the detection from features, sizes, and areas of microaneurysms. Our proposed patching sizes are 256×256, 384×448, and 768×896 that shown in Fig. 2.

3.5 UNet

UNet is an architecture that is extensively used for object segmentation, especially in the health area. Low level detail information and high level semantic information are combining by Unet, this achieves properly performance on biomedical
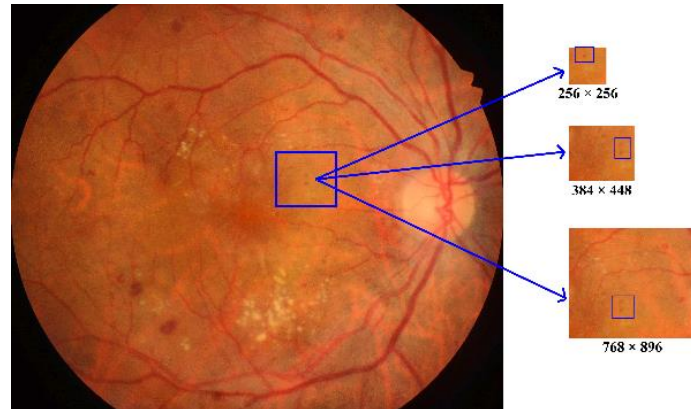
Figure. 2 Multi size patching IDRiD dataset

imagesegmentation. When the training data limited, training such a deep network is very hard. One way is applying data augmentation, as done in UNet architecture. Besides data augmentation, the architecture of UNet also contributes to decreasing the training problem. [21,31]. UNet both facilitating backward propagation through training, and compensating low level finer details to high level semantic features. This has a similar purpose to that of residual neural network [22,31]. We note that replacing a plain unit with a residual unit can improve performance on the Unet architecture.

3.6 Residual network

Using a deeper neural network architecture can improve the performance of a deep learning network, but it can hamper the training process and problems with overfitting or degradation. To overcome these problems, a deep Residual network proposed by He et al. to overcome the problem of overfitting and degradation during the training process [22,31]. The deep Residual network has a repeating Residual network layer. The Residual network y is obtained by adding the result of the identity mapping function $h(x)$ with the Residual unit function $F(x)$. Here x_l is the feature input of the l-th unit. W_l is the set weight and bias associated with the l-th unit. x_{l+1} represents the output of the l-th unit after calculating the activation function $f(y)$ using the relu [32].

$$y_l = h(x_l) + F(x_l, W_l) \qquad (5)$$

$$x_{l+1} = f(y_l) \qquad (6)$$

Residual UNet has an architecture that has fast computation time due to the identity skip connection. Fig. 3 shows the Residual units of the original Residual units and the proposed using pre-activation

Residual units. The original Residual unit consists of layers in the order of convolution, batch normalization, and relu, repeated at the next layer except relu activation is used after addition. The original Residual unit has difficulty reducing losses during training. Identity mapping with a 1×1 convolutional shortcut layer will experience a decrease in the accuracy value if the number of Residual layer stacks gets deeper. He et al. proposed a Residual unit using pre-activation with a combination of batch normalization (BN) and relu as pre-activation followed by a convolutional layer. When Batch Normalization and relu are used before the convolutional layer it can overcome over-segmentation in the edge area. For identity mapping, we propose a convolutional layer followed by batch normalization to normalize the input features of the convolutional layer and improve the performance of architecture.

3.7 MResUNet

We propose UNet-Residual because, UNet architecture can train networks with limited training data, and Residual units can train networks with fast computing time. Identity mapping on the Residuals can reduce over-segmentation in the edge area [31]. Identity mapping on the Residual unit can reduce the number of training parameters, thus allowing training with few parameters but resulting in appropriate segmentation. Fig. 3(a) is a plain UNet, Fig. 3(b) is an original Residual-identity mapping, which uses the BN after addition [32] and Fig. 3(c) is a Residual unit modification that is useful for increasing the number of training features. The architecture has 6 types of full pre-activation Residual units. The architecture consists of contracting, bridge, and expansive paths. The contracting path consists of a BN after addition and 2 full pre-activation Residual

Table 2. Details of the convolutional layers of the network architecture.

	Unit Level	Conv Layer	Filter and Stride	Output Size
Input				256×256×3
Con-tracting	Blok1	Conv1	3×3 /1	256×256×64
		Conv2	3×3 /1	256×256×64
		Conv3	3×3 /1	256×256×64
	Blok2	Conv4	3×3 /2	128×128×128
		Conv5	3×3 /1	128×128×128
		Conv6	3×3 /2	128×128×128
	Blok3	Conv7	3×3 /2	64×64×256
		Conv8	3×3 /1	64×64×256
		Conv9	3×3 /2	64×64×256
Bridge	Blok4	Conv10	3×3 /2	32×32×512
		Conv11	3×3 /1	32×32×512
Expan-sive	Blok5	Conv12	3×3 /1	64×64×256
		Conv13	3×3 /1	64×64×256
		Conv14	3×3 /1	64×64×256
	Blok6	Conv15	3×3 /1	128×128×128
		Conv16	3×3 /1	128×128×128
		Conv17	3×3 /1	128×128×128
	Blok7	Conv18	3×3 /1	256×256×64
		Conv19	3×3 /1	256×256×64
		Conv20	3×3 /1	256×256×64
Output		Conv21	3×3 /1	256×256×1

units with identity mapping to perform feature extraction and reduce the size of the input image. The bridge path consists of full pre-activation and identity mapping to connect the contracting path and the expansive path. The expansive path consists of upsampling, concatenate, and 3 full pre-activations to restore the image to its original size. Concatenate functions to connect the features of the contracting path shown in Fig. 4.

The Residual unit on the contracting path and the bridge path has a convolution with stride 2 to reduce the feature map by half. The expansive path has 3 Residual units with a stride size 1 in the convolution layer. With the last layer, the softmax layer determines the segmentation between microaneurysms and background. The total number of layers consists of 68 layers with 21 convolutional layers, the details of the convolutional layers are described in Table 2.

In the training process patches with sizes 256×256, 384×448, and 768×896 are converted to the same size 256×256 when they are become input image on the MResUnet network.

3.8 Mean Weighted Loss Function

The number of background pixels and microaneurysm pixels has a very high ratio, where the number of background pixels is the majority compared to microaneurysm pixels. Therefore cross-

entropy cannot overcome an imbalanced pixel ratio. To overcome this problem, a weighted cross-entropy loss function is proposed to create a high error when the background class appears.

N denotes the number of pixels of the image and K is the number of classes with k class representing the image, $k = \{1, 2, ..., K\}$. Where $K = 2$ consist
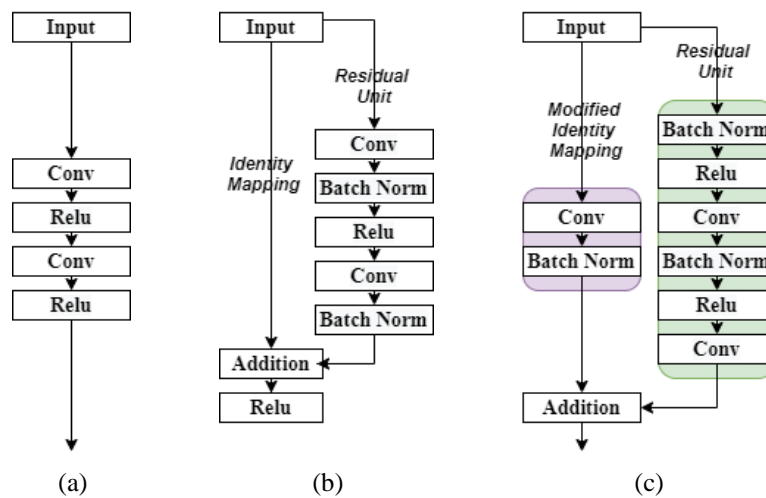

Figure. 3 Building blocks of Residual-Unet: (a) Plain unit in UNet, (b) residual unit with identity mapping, and (c) unit in Modified Residual UNet (MResUNet)
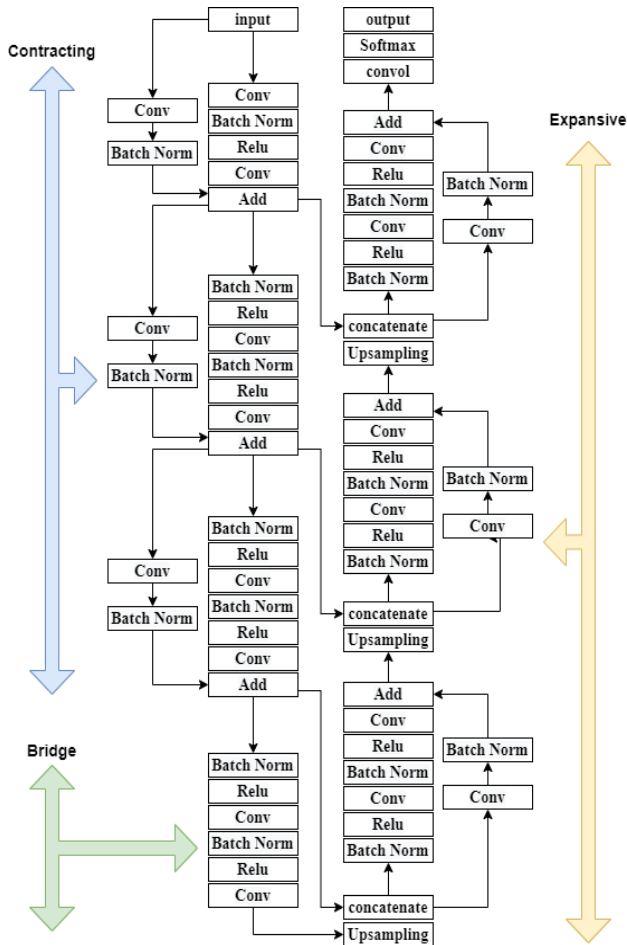
Figure. 4 Proposed MResUNet architecture

two classes in binary segmentation task. $\hat{z}_k^{(i)}$ is the actual probability of pixel $(\hat{z}^{(i)})$ that belong to k-th class, and $z_k^{(i)}$ represent the output probability of pixel $(z^{(i)})$ that belonging k-th class. Binary cross-entropy loss function can be calculated by Eq. (7).

$$CE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\hat{z}_k^{(i)} \, log\left(z_k^{(i)}\right)\right) \qquad (7)$$

The proposed weight can be applied to the binary cross-entropy loss function as in Eq. (8). In Eq. (9), the weight W_{c_k} is obtained by calculating the mean value of all pixel frequencies of all classes in the training image, $\bar{f}$, divided by the frequency itself f_{c_k} in each c_k class. The pixel frequency of each class can be calculated using Eq. (10), where n_{c_k} is the number of pixels of each class c_k.

$$WCE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(W \, \hat{z}_k^{(i)} \, log\left(z_k^{(i)}\right)\right) \qquad (8)$$

$$W_{c_k} = \bar{f}/f_{c_k} \qquad (9)$$

$$f_{c_k} = n_{c_k}/N \qquad (10)$$

3.9 Post Processing

We use post-processing to improve the quality of the segmentation results, we apply mathematical morphology to remove noise artifacts [33].

4. Experiments and analysis

The proposed research uses the IDRID dataset as training and testing [23] and DiaretDB1 as benchmark testing [10]. The dataset consists of 45 images as training data and microaneurysm ground truth. Before training, we enhanced an image using a cumulative histogram equalization to increase the contrast. Then, we cropped the retinal area to reduce the background area. The result of cropped image is shown in Fig. 5 (d). After cropping the whole image, we resize the cropped image to 3584×3072. Each resized image is patched using three different sizes, they are 256×256, 384×448, and 768×896. Different patching sizes can increase the accuracy of the network to detect the features, size, and area of microaneurysms. We use the subimages of the patches containing microaneurysms for training. The subimage dataset is divided into 90% for the train set and 10% for the validation set used for monitoring training. To conduct training using a Residual network, we apply a learning rate of 0.05 and optimize the loss using the SGDM optimizer with a momentum of 0.9. Mini batch is set to 1 and 50 epochs. The mean weighted cross entropy is applied for optimizing the cross entropy weights on the training image dataset. The mean weighted cross entropy obtained will give a high weight value for microaneurysms compared to the background. Testing images consist of 10 IDRiD images and 10 DiaretDB images for benchmarking. Each testing image is patched with a size of 1792×1536 and after the segmentation process, we reconstruct the segmentation image become the size of 3584×3072. The testing input, corresponding ground truth, and segmentation result are shown in Fig. 5 (a) , (e), and (f). Table 3 shows the dataset division.

4.1 Evaluation metrics

To evaluate the performance of the network architecture for microaneurysm segmentation, we utilize 3 metrics, they are accuracy (AC), sensitivity (SE), and specificity (SPE), based on true positive

Table 3. Dataset division

Dataset	Microaneurysms	
	Images	Patches
IDRiD training	54	5875
IDRiD validation		653
IDRiD testing	10	40
DiaretDB1 testing	10	40

Table 4. The evaluation results of the baseline networks on the IDRiD dataset.

Method	Averaged Performance (%)		
	Accuracy	Sensitivity	Specificity
Auto-Encoder [20]	**99.81**	39.52	**99.88**
FCN16 [19]	98.59	45.52	98.65
FCN8 [19]	98.09	43.18	98.15
UNet [21]	99.15	39.61	99.23
Proposed MResUNet	99.76	**61.96**	99.81

(TP), true negative (TN), false positive (FP), and false negative (FN). The performance measurement is given in Eq. (11), Eq. (12), and Eq. (13).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$sensitivity = \frac{TP}{TP + FN} \quad (12)$$

$$specificity = \frac{TN}{TN + FP} \quad (13)$$

4.2 Comparison with baseline models

In this section, we compare the performance of our proposed architecture with the baseline network, such as AutoEncoder from VGG, FCN, and UNet. The results of the quantitative comparisons are shown in Table 4 and the results of the segmentation visualization of the different methods are shown in Fig. 6.

In Table 4 the best performance on IDRiD dataset is highlighted in bold. The proposed method has the best sensitivity value compared to all methods. The proposed method is still quite good with the second accuracy and specificity after the autoencoder. This shows that MResUNet can solve imbalanced cases better than other networks. Autoencoder has a slightly higher accuracy and specificity value than MResUNet, but the autoencoder architecture is less effective in detecting microaneurysm lesions as indicated by the low sensitivity value. FCN 32 cannot detect microaneurysm because FCN 32 returns the entire image in the upsampling process. not have a sufficient spatial resolution in the direct output, which causes the architecture cannot detect microaneurysm properly.

Meanwhile, the improved architecture of FCN 32, FCN 16, and FCN 8 can detect microaneurysms FCN 32 does because FCN 16 and FCN 8 contain a 2× upsampling block that the training image is not returned to its original size at once at the upsampling process. This makes FCN16 and FCN 8 can detect microaneurysm better than FCN32.

Fig. 6 presents the visualization results of microaneurysm segmentation using the same test image. The autoencoder cannot detect the microaneurysms that shown in the yellow squares of Fig 6 (f), because they are closer to the blood vessels. The modified methods of FCN 32 that FCN 16 and FCN 8 can detect in the area given by the yellow square, but there is false positive where the blood vessels are detected as microaneurysms, shown in Fig. 6 (c) and Fig. 6 (d). The UNet architecture can detect in the area given by the yellow square, as shown in Fig. 6 (e), but UNet architecture causes the false positive of areas that are not microaneurysms.

In table 5 the best performance on the DiaretDB1 dataset is marked in bold. The proposed method has the best sensitivity and accuracy values compared to all methods. The proposed method is quite good enough in specificity value after autoencoder, but autoencoder is not good enough in detecting microaneurysms in DiaretDB1 dataset. The UNet architecture still can detect microaneurysm with the sensitivity value good enough compared with FCN 16. FCN 8 which is a modified architecture of FCN 16 can achieve better sensitivity compared to autoencoder, number two after MResUNet. This shows that MResUNet can solve the imbalance between microaneurysms and background well compared to other networks. The performance sensitivity value on the DiaretDB1 dataset is higher than the IDRiD dataset because DiaretBD1 measures its performance using ROI from the ground truth. The DiaretDB1 dataset uses ground truth red small dots instead of microaneurysm that which advances the higher performance of sensitivity value. Testing image and the corresponding ground truth are shown in Fig. 7(a) and Fig. 7(b). The results of segmentation using FCN8 and auto-encoder can detect microaneurysm one of the two microaneurysm areas marked with a yellow square that is shown in Fig. 7 (d) and Fig. 7 (f) respectively. Meanwhile, Unet and
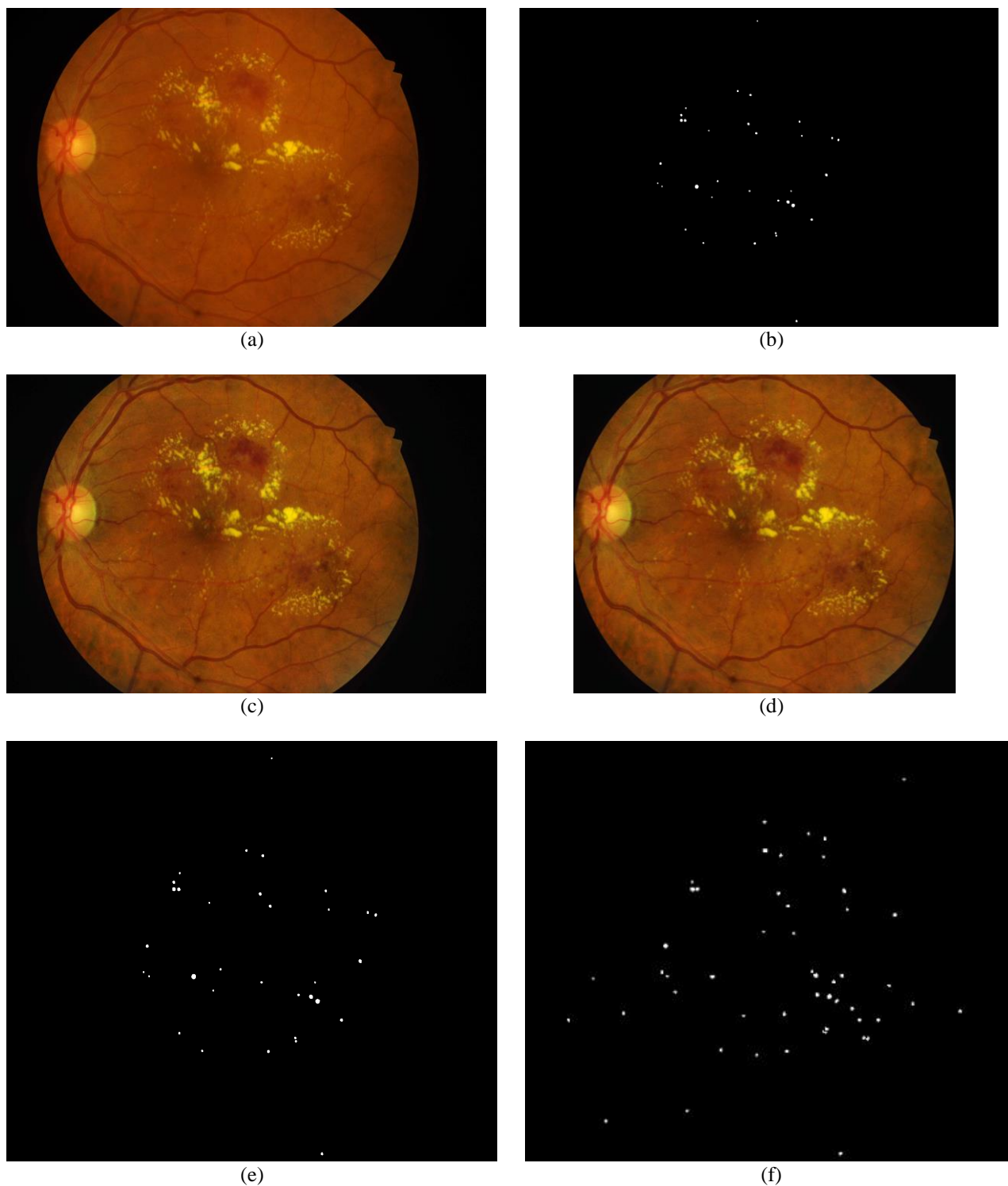
Figure. 5 Visualization of the microaneurysm segmentation results of an IDRiD image: (a) an original image, (b) the corresponding ground truth, (c) the enhancement result, (d) the cropped image, (e) corresponding ground truth after cropping, and (d) microaneurysm segmentation result

FCN16 miss in detecting microaneurysm shown in Fig. 7 (c) and Fig. 7 (e) respectively. The MResUnet can detect all microaneurysm shown in Fig. 7 (g) the yellow square, this shows that MResUnet can perform segmentation properly.

We also present a comparison chart of the maximum and minimum values in sensitivity.
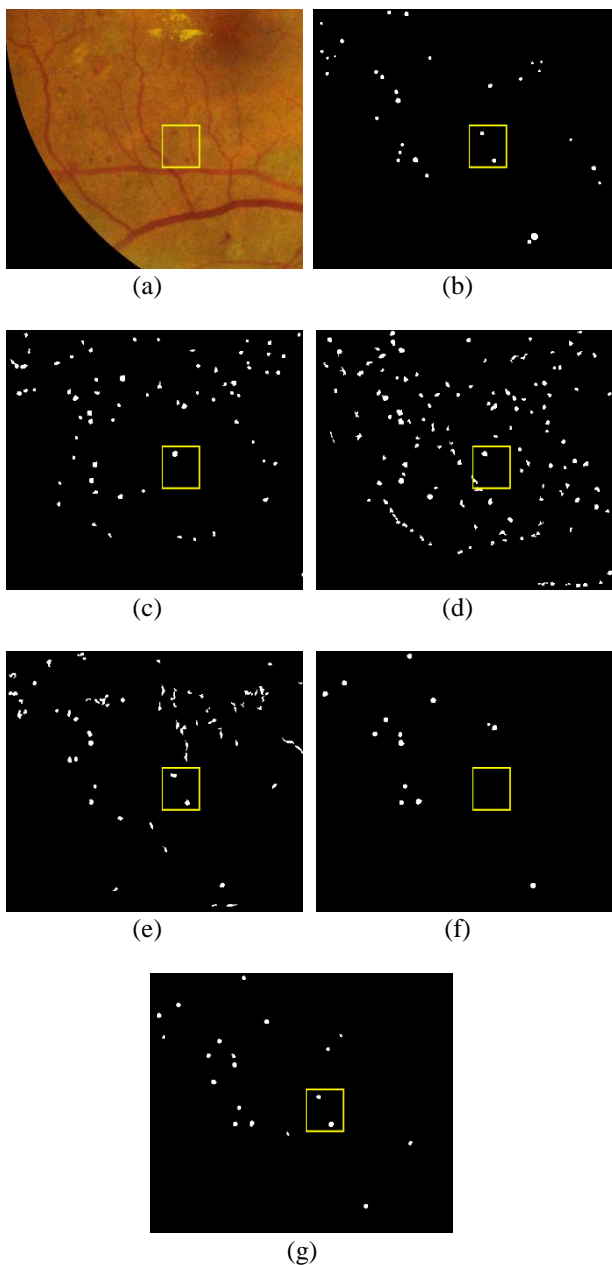
Figure. 6 Visualization results of microaneurysm segmentation on a testing image from IDRiD: (a) original test image, (b) corresponding ground truth, (c) segmentation result from FCN 16, (d) Segmentation result from FCN 8, (e) segmentation result from UNet, (f) segmentation result from auto-encoder, and (g) segmentation result from our proposed method using MResUNet
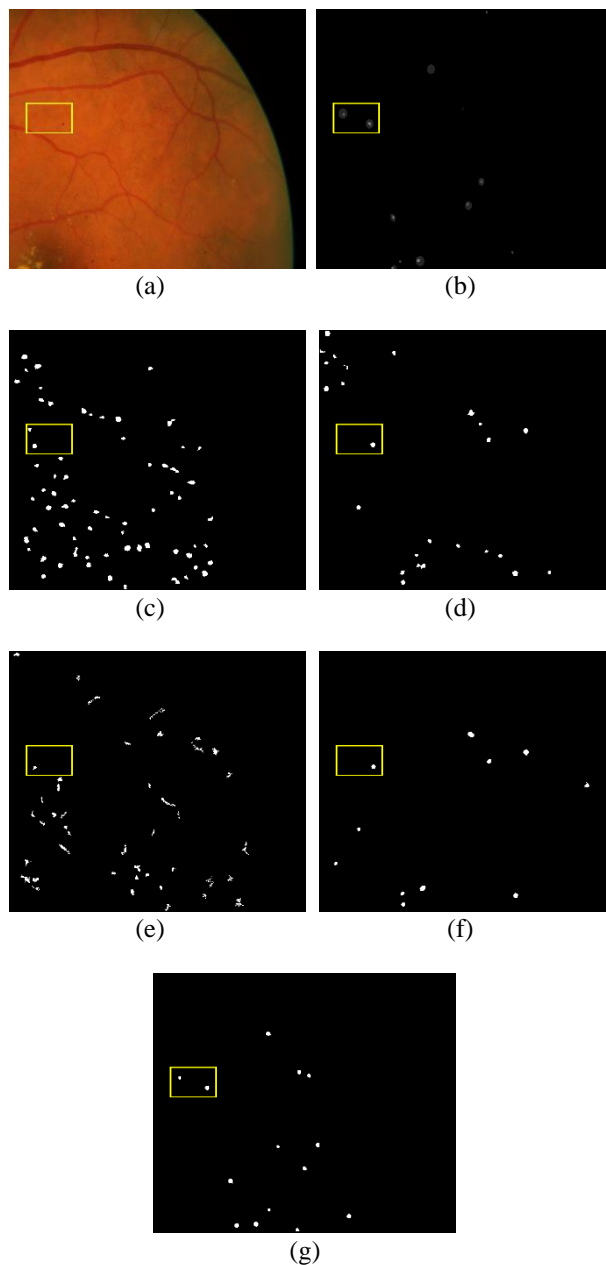
Figure. 7 Visualization results of microaneurysm segmentation on a testing image from DiaretDB1: (a) original test image, (b) corresponding ground truth, (c) segmentation result from FCN 16, (d) Segmentation result from FCN 8, (e) segmentation result from UNet, (f) segmentation result from auto-encoder, and (g) segmentation result from our proposed method using MResUNet

As shown in Fig. 8, the maximum and minimum value of the proposed method, MResUNet, is better compared with the other architectures, and achieves a value of 71.33% and 53.80% respectively, which indicates that the proposed method is sensitive to detect the microaneurysm areas properly.

Sarhan et al. [18] proposed two stage deep learning with multi-scale input and triplet loss for microaneurysm segmentation. Sarhan et al. applied the two-step deep learning hypothesis generation network (HGN) and patch-wise refinement network (PRN). Where PRN is an advanced network to improve results after the HGN process.

Comparison performance between the proposed method and the state of the art method, using average sensitivity. The method proposed by Sarhan et al. [18] won 4th place in the IDRiD competition with the sensitivity value of 46.26%. The proposed method has higher sensitivity with the value of 61.96%. It

shows that the proposed method is architecturally sensitive for detecting microaneurysms.

4.3 Evaluation of loss function

The performance of the MResUNet architecture using a different loss function is compared, shown in Table 6. The loss functions being compared are cross-entropy, Dice, Tversky, and the proposed mean weight cross entropy. Experiments were applied to the IDRiD dataset.

Dice loss [34] and Tversky Loss [35] are loss functions that do not require weight values. The advantage of both methods is overcoming imbalance without adding weight to the loss function. However, the modified cross entropy using weights can handle imbalance better than Dice and Tversky loss. The result is shown in Table 6.

Based on the performance evaluation results in Table 6, it shows that using the mean weight on the cross entropy loss function achieves high sensitivity values, especially in datasets with very high imbalanced classes such as microaneurysms in the IDRiD dataset. A very high imbalanced class obtains a low sensitivity value for the loss function that does not use weights. The use of weights in the loss function increases the high probability value in the minority class. The results of the segmentation using different loss functions for the MResUNet architecture are shown in Fig. 8.

Based on Fig. 9 (c), Fig. 9 (d), and Fig. 9 (e) the cross entropy, Dice, and Tversky loss function can detect a few areas of microaneurysms. In Fig. 9 (f) the mean weighted cross entropy, can detect the microaneurysms area is better than cross entropy without weight, Dice, and Tversky, shown with the yellow square at the top. Cross entropy, Dice, and Tversky cannot detect microaneurysms because of the area close to blood vessels, while mean weighted can detect microaneurysms properly.

We evaluate the misclassification of our proposed method by making visualization of the sample segmentation result for a testing image. In Fig. 10, the segmentation result is presented. The white color indicates true positive, cyan is false negative, and red indicates false positive. Based on the image, many false positives were detected in areas close to blood vessels, and the area around hemorrhages which has the same intensity as microaneurysms.

Table 5. The evaluation results of the baseline networks on the DiaretDB1 dataset.

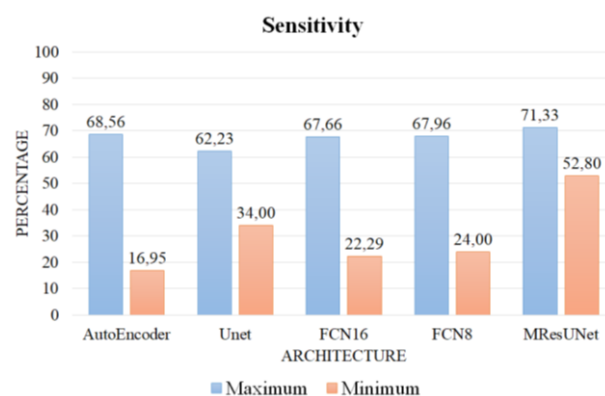| Method	Averaged Performance (%)		
	Accuracy	Sensitivity	Specificity
Auto-Encoder [20]	99.71	43.53	**99.80**
FCN16 [19]	98.00	11.39	98.15
FCN8 [19]	99.55	65.40	99.60
UNet [21]	99.13	22.40	99.26
Proposed MResUNet	**99.75**	**85.87**	99.77

Figure. 8 Comparison of sensitivity of different architectures on IDRiD dataset

Table 6. Comparison of the loss function performances for the proposed MResUNet

Loss Function	Averaged Performance (%)		
	Accuracy	Sensitivity	Specificity
Cross entropy	99.87	22.89	99.98
Dice [34]	99.88	15.97	99.99
Tversky [35]	99.87	23.04	99.98
Mean Cross Entropy	99.76	**61.96**	99.81

False negatives occur in areas that are attached to blood vessels and hemorrhages. False negatives also occur in areas with low contrast, so that the network loses information to detect these areas as microaneurysms. Based on the example, the upper squares show the false negatives that occur in the area attached to the hemorrhages so that the microaneurysms are misclassified by the network, because of the similarity in the intensity value with the hemorrhages. False negatives also occur in areas that have dark intensity and irregular size so that they are thought to be hemorrhages. In the bottom left square of the example, the false positive is dark and rounded around the blood vessel, so that it is falsely predicted as a microaneurysm.

Based on Fig. 9 (c), Fig. 9 (d), and Fig. 9 (e) the cross entropy, Dice, and Tversky loss function can detect a few areas of microaneurysms. In Fig. 9 (f) the mean weighted cross entropy, can detect the microaneurysms area is better than cross entropy without weight, Dice, and Tversky, shown with the yellow square at the top. Cross entropy, Dice, and Tversky cannot detect microaneurysms because of the area close to blood vessels, while mean weighted can detect microaneurysms properly.

We evaluate the misclassification of our proposed method by making visualization of the sample segmentation result for a testing image. In Fig. 10, the segmentation result is presented. The white color indicates true positive, cyan is false negative, and red indicates false positive. Based on the image, many false positives were detected in areas close to blood vessels, and the area around hemorrhages which has the same intensity as microaneurysms.

False negatives occur in areas that are attached to blood vessels and hemorrhages. False negatives also occur in areas with low contrast, so that the network loses information to detect these areas as microaneurysms. Based on the example, the upper squares show the false negatives that occur in the area attached to the hemorrhages so that the microaneurysms are misclassified by the network, because of the similarity in the intensity value with the hemorrhages. False negatives also occur in areas that have dark intensity and irregular size so that they are thought to be hemorrhages. In the bottom left square of the example, the false positive is dark and rounded around the blood vessel, so that it is falsely predicted as a microaneurysm.

5. Conclusion

Microaneurysm segmentation is still a challenge for researchers at this time. In this paper, we propose MResUNet to detect microaneurysms by applying the skip connection in UNet, to get proper segmentation. In addition, we also propose adaptive weighted loss function cross-entropy on the MResUNet to overcome the problem of pixel imbalance between microaneurysms and background. The weight values are calculated based on the mean value of the frequencies of image training pixels. The proposed weighted loss function can improve the network architecture to detect the microaneurysm class which is a minority class and can apply to multiclass segmentation.

We conducted an architecture evaluation using the IDRiD and DiaretDB1 dataset as a benchmark test. The experimental results show that the proposed MResUNet architecture is good enough for

sensitivity compared to the state of the art method. The MResUNet network achieved higher sensitivity

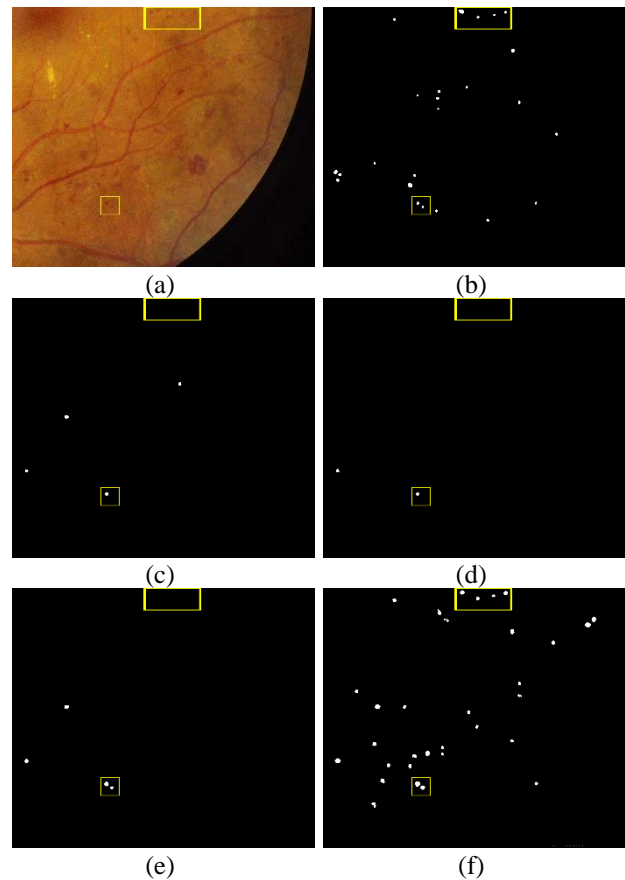

(a) (b)

(c) (d)

(e) (f)

Figure. 9 Visualization of microaneurysm segmentation on IDRiD testing data using different loss functions: (a) original test image, (b) corresponding ground truth, (c) segmentation result using Cross Entropy, (d) segmentation result using Dice, (e) segmentation result using Tversky, (f) segmentation result using Mean Cross Entropy
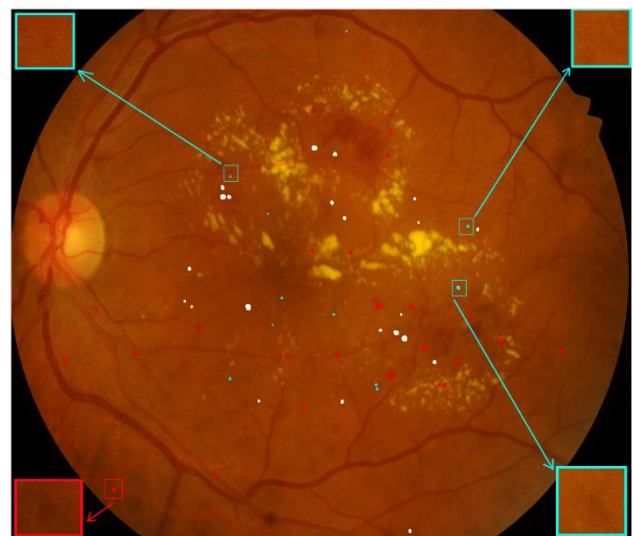

Figure. 10 The example result of microaneurysm segmentation in a IDRiD image

values 61.96% and 85.87% on the IDRID and DiaretDB1 datasets, respectively, compared to the AutoEncoder, FCN16, FCN8, and UNet architectures. Comparison of the loss function on the MResUnet architecture using the Mean Cross

Entropy resulted in a higher sensitivity value of 61.96% compared to Cross Entropy, Dice, Tversky loss. Cross entropy weighting with the mean value is more sensitive to detect microaneurysm areas. However, there is still a problem of misclassification in images that have low contrast. It needs further research to improve the performance value.

Our future work of this research is to develop segmentation systems of exudates and hemorrhages.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Dinial Utami Nurul Qomariah contributed to the formulation of the proposed method, deep learning architecture, the experiments, and the writing of the original draft preparation. Handayani Tjandrasa supervised in the conceptualization, the problem formulation, the conduction of the reseach process, and article preparation. Chastine Fatichah supervised the problem formulation, the conduction of the reseach process, and article preparation.

References

[1] N. B. A. Mustafa, W. M. D. W. Zaki, and A. Hussain, "A review on the diabetic retinopathy assessment based on retinal vascular tortuosity", In: *Proc. of International Colloquium on Signal Processing Its Applications (CSPA)*, pp. 127–130, 2015.

[2] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030", *Diabetes Care*, Vol. 27, No. 5, pp. 1047–1053, May 2004.

[3] R. Putra, H. Tjandrasa, and N. Suciati, "Severity Classification of Non-Proliferative Diabetic Retinopathy Using Convolutional Support Vector Machine", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, pp. 156–170, 2020.

[4] C. P. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J. T. Verdaguer, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales", *Ophthalmology*, Vol.

110, No. 9, pp. 1677–1682, 2003.

[5] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM", In: *Proc. of International Conference on Information Communication Technology and System (ICTS)*, pp. 152–157, 2019.

[6] G. S., V. P. Gopi, and P. Palanisamy, "A lightweight CNN for Diabetic Retinopathy classification from fundus images", *Biomedical Signal Processing and Control*, Vol. 62, p. 102115, 2020.

[7] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model", *Pattern Recognition Letters*, Vol. 133, pp. 210–216, 2020.

[8] B. Al-Diri, F. Calivá, P. Chudzik, G. Ometto, and M. Habib, "Chapter 12 - Diabetic retinopathy and maculopathy lesions", in *The Elsevier and MICCAI Society Book Series*, E. Trucco, T. MacGillivray, and Y. B. T.-C. R. I. A. Xu, Eds. Academic Press, pp. 223–243, 2019.

[9] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy", *Pattern Recognition*, Vol. 46, No. 1, pp. 107–116, 2013.

[10] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "DIARETDB1 diabetic retinopathy database and evaluation protocol", In: *Proc. Medical Image Understanding and Analysis (MIUA)*, pp. 1–18, 2007.

[11] J. Xue, S. Yan, J. Qu, F. Qi, C. Qiu, H. Zhang, M. Chen, T. Liu, D. Li, and X. Liu, "Deep membrane systems for multitask segmentation in diabetic retinopathy", *Knowledge-Based Systems*, Vol. 183, p. 104887, 2019.

[12] P. Chudzik, S. Majumdar, F. Calivá, B. Al-Diri, and A. Hunter, "Microaneurysm detection using fully convolutional neural networks", *Computer Methods and Programs in Biomedicine*, Vol. 158, pp. 185–192, 2018.

[13] A. Manjaramkar, and M. Kokare, "Connected Component Clustering Based Hemorrhage Detection in Color Fundus Images", *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 2, pp. 143–151, 2018.

[14] H. Jaafar, H. Al-Libawy, and Q. Al-Gayem, "Automated Approach for Extraction of

Microaneurysms and Hemorrhages in Retinal Fundus Images", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 5, pp. 226–237, 2020.

[15] M. Haloi, "Improved Microaneurysm Detection using Deep Neural Networks", *arXiv*, 2015.

[16] C. Lam, C. Yu, L. Huang, and D. Rubin, "Retinal Lesion Detection With Deep Learning Using Image Patches.", *Investigative Ophthalmology & Visual Science*, Vol. 59, No. 1, pp. 590–596, 2018.

[17] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "An ensemble deep learning based approach for red lesion detection in fundus images", *Computer Methods and Programs in Biomedicine*, Vol. 153, pp. 115–127, 2018.

[18] M. H. Sarhan, M. Yigitsoy, and A. Eslami, "Multi-scale Microaneurysms Segmentation Using Embedding Triplet Loss", In: *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 174–182, 2019.

[19] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, p. 1, 2016.

[20] H. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1930–1943, 2013.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, pp. 234–241, 2015.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", In: *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[23] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research", *Data*, Vol. 3, No. 3, 2018.

[24] P. Khojasteh, B. Aliahmad, and D. K. Kumar, "Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms", *BMC ophthalmology*, Vol. 18, No. 1, p. 288, Nov. 2018.

[25] J. H. Tan, H. Fujita, S. Sivaprasad, S. V Bhandary, A. K. Rao, K. C. Chua, and U. R. Acharya, "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network", *Information Sciences*, Vol. 420, pp. 66–76, 2017.

[26] S. Guo, T. Li, H. Kang, N. Li, Y. Zhang, and K. Wang, "L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images", *Neurocomputing*, Vol. 349, pp. 52–63, 2019.

[27] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system", In: *Proc. of International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2392–2397, 2014.

[28] P. Garg, and T. Jain, "A Comparative Study on Histogram Equalization and Cumulative Histogram Equalization", *International Journal of New Technology and Research (IJNTR)*, Vol. 3, No. 9, pp. 41–43, 2017.

[29] S. Jenifer, S. Parasuraman, and A. Kadirvelu, "Contrast enhancement and brightness preserving of digital mammograms using fuzzy clipped contrast-limited adaptive histogram equalization algorithm", *Applied Soft Computing*, Vol. 42, pp. 167–177, 2016.

[30] A. M. Reza, "Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement", *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, Vol. 38, No. 1, pp. 35–44, 2004.

[31] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net", *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, pp. 749–753, 2018.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks", In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 630–645, 2016.

[33] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2003.

[34] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations", *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2017, pp. 240–248.

[35] S. S. Salehi, D. Erdogmus, and A. Gholipour, "Tversky Loss Function for Image

Segmentation Using 3D Fully Convolutional Deep Networks", in *arXiv*, pp. 379–387, 2017.