



Residual Fully Convolutional Network for Mandibular Canal Segmentation

Yolla Faradhilla^{1*} Agus Zainal Arifin¹ Nanik Suciati¹ Eha Renwi Astuti²
 Rarasmaya Indraswari³ Monica Widiarsi¹

¹Department of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia

²Department of Dentomaxillofacial Radiology, Universitas Airlangga, Indonesia

³Department of Information Systems, Institut Teknologi Sepuluh Nopember, Indonesia

* Corresponding author's Email: yollafrdl@gmail.com

Abstract: In performing dental implant surgery on the mandible, it is necessary to measure the height and width between alveolar bone and mandibular canal. A system that is able to measure the height and width automatically is required since a manual measurement takes a long time. One of the important processes in this system is mandibular canal segmentation. The main problem in this segmentation process is the unbalanced numbers of data between object and background classes. This problem often led to misclassifications, especially at pixels on the object boundary. This research proposed a new architecture based on residual fully convolutional network (RFCN) by considering the loss values in the region and boundary of the mandibular canal segmentation. Dual auxiliary loss (DAL) functions are introduced to optimize RFCN, so that the network performs object segmentation better. For this research, DAL used focal loss to calculate the loss value in region and boundary to overcome the unbalanced class between the mandibular canal and the background. There are 2 datasets used for this research. The first dataset contains 200 images with mandibular canal and the second dataset contains 300 images consisting of 200 images with mandibular canal and 100 images without mandibular canal. We tested our network and compared it with state-of-the-art segmentation methods. The experiment showed that the proposed method outperforms all the comparing methods with Dice Similarity Score of 0.914 on the first dataset and 0.868 on the second dataset.

Keywords: Mandibular canal segmentation, Fully convolutional network, Residual network.

1. Introduction

In general, adult humans have 32 teeth for tearing and chewing food [1]. In addition, teeth can also affect human articulation in speech. With these functions, tooth loss will be a problem for humans. Nowadays, if someone loses a tooth, they can perform dental implant surgery. The success of dental implant surgery has been validated and considered positive by the public [2]. Dental implants are artificial tooth roots that will be implanted in the human jaw to replace the roots of the missing teeth. While placing dental implants, many things need to be considered, especially the mandibular canal and maxillary sinus to avoid the injury of inferior alveolar nerve and maxillary sinus [3].

Mandibular implants are implants that are placed in the human mandible. There are 3 things that must

be considered before placing an implant in the mandible, namely bone density, distance of height and width between the alveolar bone and the mandibular canal, and the distance between the teeth or implants [3]. Measurement of the height and width between the alveolar bone and the mandibular canal takes a long time if done manually. For that, it is necessary to have a system that can measure these measurements automatically. One of the important processes for measuring the height and width distance between the alveolar bone and the mandibular canal automatically is mandibular canal segmentation. The mandibular canal or inferior alveolar canal is located in the lower jaw which has an S-shaped pattern with a diameter between 2 – 2.4 mm [4].

Many researches have been done to segment the mandibular canal. Moris *et al.* [5] tracked the

mandibular canal automatically using the matching method and multiple hypotheses. Several candidate positions were selected by looking for a high contrast in mandible. Then, several candidates for the mental foramen and mandibular positions will be evaluated. However, when there was a position that was not extracted correctly, the result in the next step will have a chance to fail.

Then, Abdolali, *et al.* [6, 7] and Kainmueller *et al.* [8] used statistical shape models (SSM) for automated segmentation of mandibular canal. The shape produced by SSM is usually biased against the reference form. However, the segmentation process was need to do 2 times, the segmentation of mandibular bone and the segmentation of mandibular canal. The segmentation of mandibular canal was done after the segmentation of mandibular bone process completed. To do 2 different segmentation processes, of course lots of ground truths that must be made.

Currently, deep learning is widely used to solve classification and segmentation problems in the health sector [9], one of which is the segmentation of the mandibular canal. Kwak *et al.* [10] tried to compared the popular deep learning method for segmentation, which is SegNet and U-Net for mandibular canal segmentation. They tried SegNet and U-Net method for 2D images and U-Net for 3D images.

U-Net [11] has been widely used to segment object, especially biomedical images. U-Net has skip connection to propagated high resolution features to the up-sampled features in decoder. Unlike U-Net, SegNet [12] used skip connection to propagated pooling indices from encoder to decoder to create sparse feature maps. SegNet used pretrained weight from basic VGGNet as encoder. The results still showed numbers of wrong pixels detected. The values of mean IoU from those study was quite low. Pixel detection error happened because of the unbalanced number of classes between object and background. The unbalanced number of classes made it possible to experience degradation information during the propagation process in skip connections on the deep network like U-Net and SegNet [13]. Residual U-Net (ResUnet) [14] was proposed to add residual network in U-Net to prevent the degradation problem. ResUnet could outperformed the other method including U-Net in road extraction from remote sensing images.

Jaskari *et al.* [9] tried to used fully convolutional neural network (FCNN) to automatically segmented the mandibular canal. The architecture of FCNN was similar with ResUnet but they used transposed convolution layer as up-sampling instead of pooling

layer. They tried to compared their FCNN method with SSM-based method like [6, 7, 8]. FCNN outperformed the other SSM-based method, but many pixels are still misclassified especially near the boundary of the object.

In the segmentation or classification process with deep learning, selecting the appropriate loss is very important because it affects the training process which can help improve model reliability [15]. In general, the loss function used by previous studies only calculates the region of the object. Two auxiliary loss was introduced [16] to assist models in segmenting portrait images in real time, namely segmenting objects (human faces) against the background. Two auxiliary loss calculates the loss value based on region and boundary. The loss function used is binary-cross entropy [17] for the region and focal loss [18] for the boundary. This calculation of loss based on region and boundary is proven to be able to classify pixels around the boundary more accurately. However, the use of binary-cross entropy can only work well on data with a balanced distribution of pixels per class [15].

To solve the problem of information degradation due to unbalanced classes, it is necessary to have a residual unit and a fully convolutional network to overcome the problem of information degradation in the model training process and also to make it easier for the model to find the right filter. To improve the model's reliability, two auxiliary losses were introduced [16]. However, the use of the binary-cross entropy loss function for region loss is not effective for data with an unbalanced class.

This research proposed an architecture based on residual fully convolutional network (RFCN) and dual auxiliary loss (DAL) for mandibular canal segmentation. DAL was introduced to optimize RFCN, which considering the loss value in the region and boundary of the mandibular canal segmentation. The region and boundary loss were merged using variable weight values and used to jointly train the RFCN model. For this research, DAL used focal loss to calculate the loss value in region and boundary to overcome the unbalanced class between the mandibular canal and the background. The proposed automatic segmentation of mandibular canal will make the process of dental implant surgery planning faster.

This paper is organized as follows. Section 2 presents the proposed methodology for mandibular canal segmentation. Section 3 presents the datasets that were used for the experiments, the experiments that have been performed, the results, and the discussion of the experimental results. Section 4 presents the conclusion and the future work of this

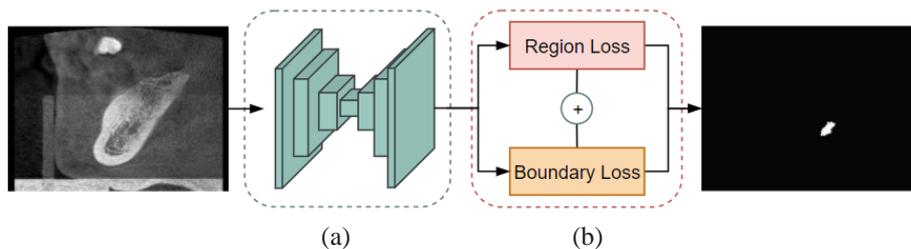


Figure. 1 Overall framework: (a) residual fully convolutional network and (b) dual auxiliary loss

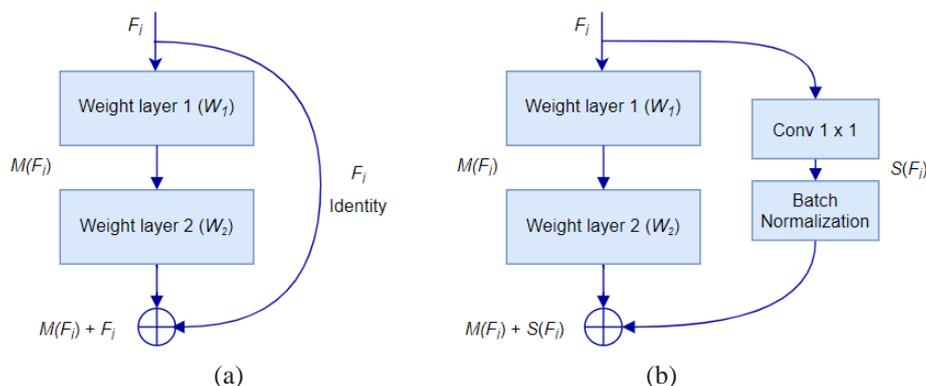


Figure. 2 Residual building block: (a) with identity mapping and (b) with 1×1 convolution layer

research.

2. Research method

The framework of the mandibular canal segmentation process is shown in Fig. 1. The residual fully convolutional network (RFCN) model in Fig. 1 (a) are used to extract features from CBCT images, while dual auxiliary loss (DAL) functions in Fig. 1 (b) are used to train the model in order to properly segment the mandibular canal. DAL calculates the loss value based on region and boundary by giving weight variables to make the calculation of loss more flexible according to the data.

2.1 Residual fully convolutional network (RFCN)

Residual fully convolutional network combine residual unit in fully convolutional network. In this research, we adopted U-Net architecture form because U-Net proven to ease the training process [14]. The 2D U-Net-like architecture was proven to work well in many medical segmentation task [15, 16]. Even with a small number of input, U-Net could accurately segment the object in a short processing time [21]. The skip connections in U-Net connected the encoder and decoder, which created an easier way to spread the information between low and high levels so that the training process will be easier [14]. We used fully convolutional network, which used convolution layer for down sampling instead of pooling layer. It can help the model on the training

process find the optimal filter and reduced the memory usage [22]. Unfortunately, fully convolutional network could cause a degradation information because of its deeper network [13].

Therefore, we combined the fully convolutional network with residual unit. This residual unit serves to prevent degradation or loss of spatial information and prevent the network from experiencing a vanishing gradient [23]. In addition, residual unit can also improve the performance of the neural network, making it easier for the training process [14]. With the addition of residual units on the fully convolutional network, fewer parameters will be used and the performance for segmenting objects will be better [14].

There were 2 kinds of residual unit or residual building block (RBB) as shown in Fig. 2, the residual building block with identity mapping and with 1×1 convolution layer. The addition of the original feature map or identity mapping mitigates the degradation issue that emerges in the models [24]. The RBB with identity mapping was used more often than RBB with 1×1 convolution layer [20, 19]. In this research, we used the RBB with 1×1 convolution layer to made the residual unit had the same number of layers as the output of weight layer for the element-wise addition. The RBB with 1×1 convolution layer was given in equation 1. The output x_{i+1} of RBB was the element-wise addition between residual function $M(F_i, W_i)$ and 1×1 convolution layer with batch normalization

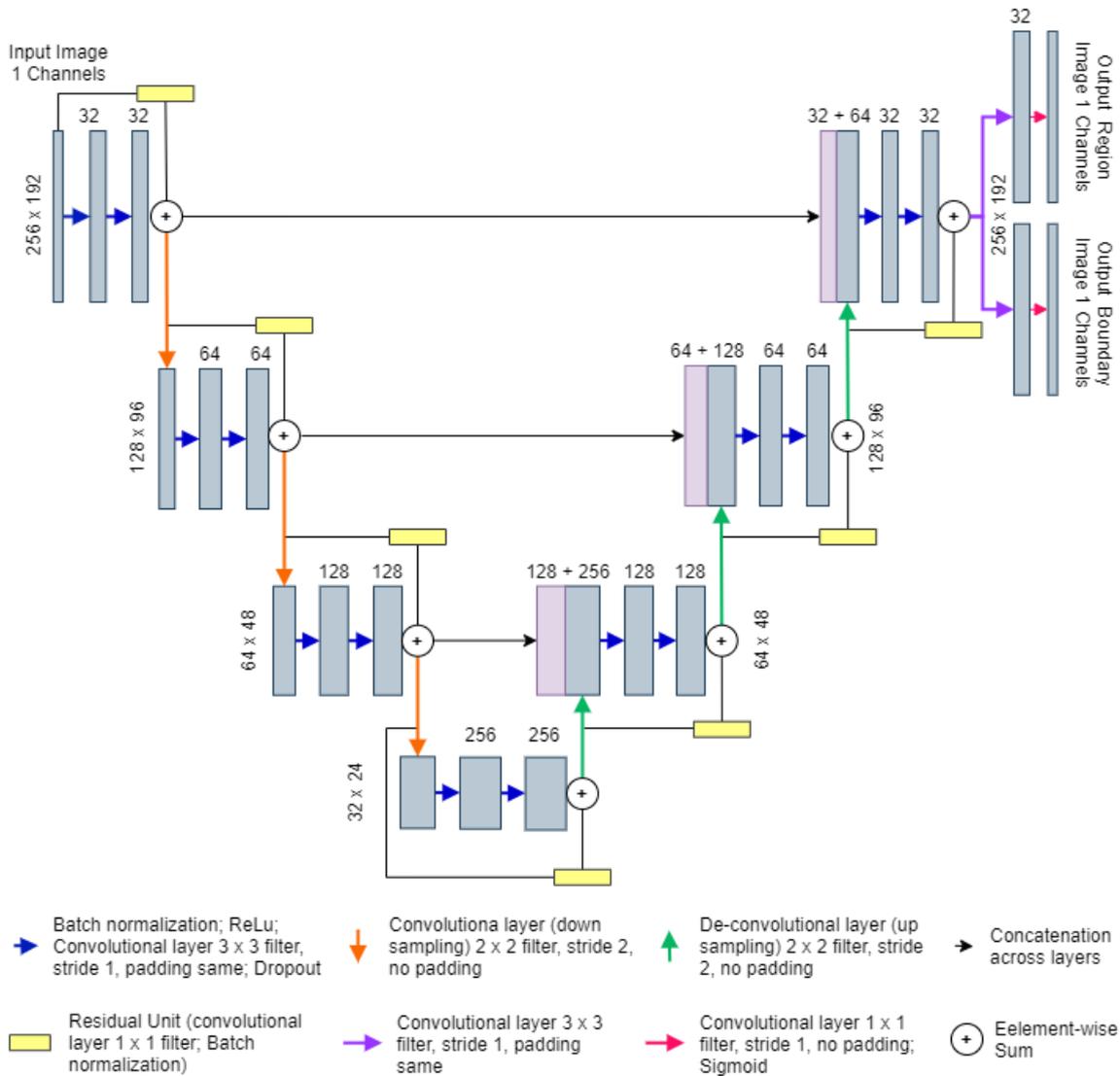


Figure. 3 Residual fully convolutional network architecture

$S(F_i)$ then multiply it with activation function f [14,20]. Residual function consisted of input F_i and the weighting process W_i like convolution layer.

$$y_i = M(F_i, W_i) + S(F_i) \tag{1}$$

$$x_{i+1} = f(y_i),$$

Fig. 3 showed the architecture used in this research. The input was 2D images with 1 channel or grayscale. The model was divided into 3 parts, there were encoder, bridge, and decoder. Each of them consisted with residual block. There were 7 residual blocks, which is 3 in encoder, 1 in bridge, and 3 in decoder. The residual block looked like Fig. 2 (b). The weight layer or convolution block consisted of batch normalization, rectified linear unit (ReLU), 3×3 convolution layer, and dropout layer with rate = 0.2. Batch normalization was used for accelerating the model’s convergence [25]. For the activation

function, we used ReLu. To ease the overfitting of the model, we used dropouts [26] as regularization function. There were 3 processes of down sampling in the encoder. We used 2×2 convolution layer with stride = 2 instead of pooling layer. Then, there were 3 processes of up sampling in the decoder. We used 2×2 transposed convolution layer with stride = 2 for up sampling. There were 2 outputs of this model, the region and boundary of mandibular canal. After the last residual blocks on decoder parts, 3×3 convolution layer was added to help the model distinguish between region and boundary. Then we added 1×1 convolutional network with sigmoid activation function to classified each of the pixels of every images.

2.2 Dual auxiliary loss (DAL)

loss function was used to help the training process so that the model can convergence quickly

[15]. In designing an architecture to carry out the segmentation process, the selection of a loss function must be adjusted to the data because it greatly influences the training process [15]. This study used dual auxiliary loss (DAL), which was the modified of two auxiliary loss [16]. DAL counted the loss value based on region and boundary by providing a variable weight (α) to made it more flexible. The DAL were donated as Eq. 2.

$$DAL = \alpha Region_{Loss} + (1 - \alpha) Boundary_{Loss} \quad (2)$$

The calculation of region loss will help the network segmented the regions of the mandibular canal while the boundary loss will help the network detect and limit objects so that it is hoped the network will not segment objects outside the boundaries. The size of the mandibular canal is very small, so the number of classes for objects and backgrounds is very unbalanced. To solve the problem of unbalanced class, then focal loss used to calculate the loss value of the region and boundary. Focal loss worked very well at data with unbalanced number of classes between object and background [13, 11]. Moreover, focal loss was proved can improve the classification's accuracy compared to cross-entropy loss [27]. Focal loss will give weight to the object which is not detected properly [15]. Focal loss was a modified of binary cross-entropy loss. The binary cross-entropy loss equation donated as equation 3. The estimated probability value for the class labelled $y = 1$ was p and to simplify the equation, focal loss used Eq. 4 to define the probability of p_t in class y [18]. So, now the cross entropy donated as Eq. 5.

$$CE_{py} = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1 - p) & \text{otherwise,} \end{cases} \quad (3)$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \quad (4)$$

$$CE_{(p,y)} = CE_{(p_t)} = -\log(p_t) \quad (5)$$

Focal loss add a modulation factor $-(1 - p_t)^\gamma$ with $\gamma \geq 0$ [18]. The value of γ would help adjusted the value for the object whose weight was lowered so that the object will get a small loss value [18]. The α used to balanced the value based on the inverse class frequency and can also be used as a hyperparameter with a value between 0 – 1 [15]. The focal loss donated as Eq. 6.

$$FL_{p_t} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (6)$$

3. Result and analysis

3.1 Dataset

In this research, we used CBCT image from 5 patients who will have a dental implant surgery as a training and testing data. The size of each CBCT image was 266 x 266 x 200 voxel. The 3D CBCT image was sliced along the sagittal plane to obtain 2D images sized 266 x 200 pixels. We randomly chose 100 images that contained the mandibular canal and 100 images that didn't contain the mandibular canal. Then, the images that showed the mandibular canal was all augmented horizontally. We trained and tested our model with 2 datasets. The first dataset contained 200 2D image that showed the mandibular canal and the second dataset contained 200 2D image that showed the mandibular canal that used in the first dataset and 100 2D image that didn't show the mandibular canal. We normalized all of the images to the intervals [0, 1]. The region and boundary of mandibular canals were annotated manually and was confirmed by experts.

3.2 Experimental setup

We implemented our model using Keras Python and Google Colaboratory GPU. To optimized the model, we used Adam with learning rate set to 0.001 and batch size set to 48. We used 1000 epochs for training the model and used model with smallest loss for testing. For dual auxiliary loss, we tried the weight possibilities from 0 – 1 with +- 0.1 for both datasets. We set the focal loss γ to 2 and α to 0.25.

3.3 Comparisons and discussions

We did 2 experiments for this research. First, we tried to find the best weight for our proposed dual auxiliary loss (DAL). Then, we compared our method with previous deep-learning-based methods for mandibular canal segmentation to validated our method. The comparison methods used are our RFCN with only considering the region loss, fully convolutional neural network (FCNN) [9], residual U-Net (ResUnet) [14], SegNet [12], and U-Net [11]. The data used for these experiments were 160 images for training and 40 images for testing in the first dataset. In the second dataset, we used 240 images and 80 images for testing. We used the score of Dice Similarity Coefficient (DSC) to evaluated the model. DSC frequently used for comparing the segmentation result and ground truth [28]. The DSC equation was donated as Eq. 7.

Table 1. Comparison of the final results

Weight		Dice Similarity Coefficient	
Region (α)	Boundary ($1 - \alpha$)	1st Dataset	2nd Dataset
1	0	0.014	0.009
0.9	0.1	0.839	0.774
0.8	0.2	0.870	0.716
0.7	0.3	0.814	0.803
0.6	0.4	0.914	0.859
0.5	0.5	0.852	0.812
0.4	0.6	0.841	0.847
0.3	0.7	0.897	0.827
0.2	0.8	0.882	0.868
0.1	0.9	0.850	0.790
0	1	0.009	0.012

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad (7)$$

TP , FP , and FN is true positive, false positive, and false negative. DSC range is from 0 – 1 where 1 indicating perfect segmentation [22].

3.3.1. Weight for dual auxiliary loss

In this experiment, we tried to find the best weight for DAL. The weight (α) value used for this experiment is between 0 – 1 for both datasets. Table 1 summarized the final result of both datasets. Final result is obtained by merging each of region and boundary output of the model. The DSC showed in Table 1 is the average DSC of the overall final result of testing data from each dataset.

In the first dataset, giving 0.6 weight for region and 0.4 for boundary gave the best DSC with 0.914 and in the second dataset, giving 0.2 weight for region and 0.8 for boundary gave the best DSC with 0.868. In both datasets, final result with the lowest DSC was when the weight of region or boundary 0. It happened because when the weight was 0, the output image was not extracted perfectly. The output image was still almost the same as the input. It will affect the final result in the merging process. The best weight for each dataset was different because each dataset has a different image characteristic. So, there was no fixed weight to use for DAL because it depends on the image condition

The samples output from the first dataset of the proposed method are shown in Fig. 4. We can see that the shape, size, orientation, and location of the mandibular canal is different for each image. Furthermore, the object around the mandibular canal, the thickness of the border, and the grayscale intensity of mandibular canal is various as well.

Mostly, the shape of mandibular canal is like an

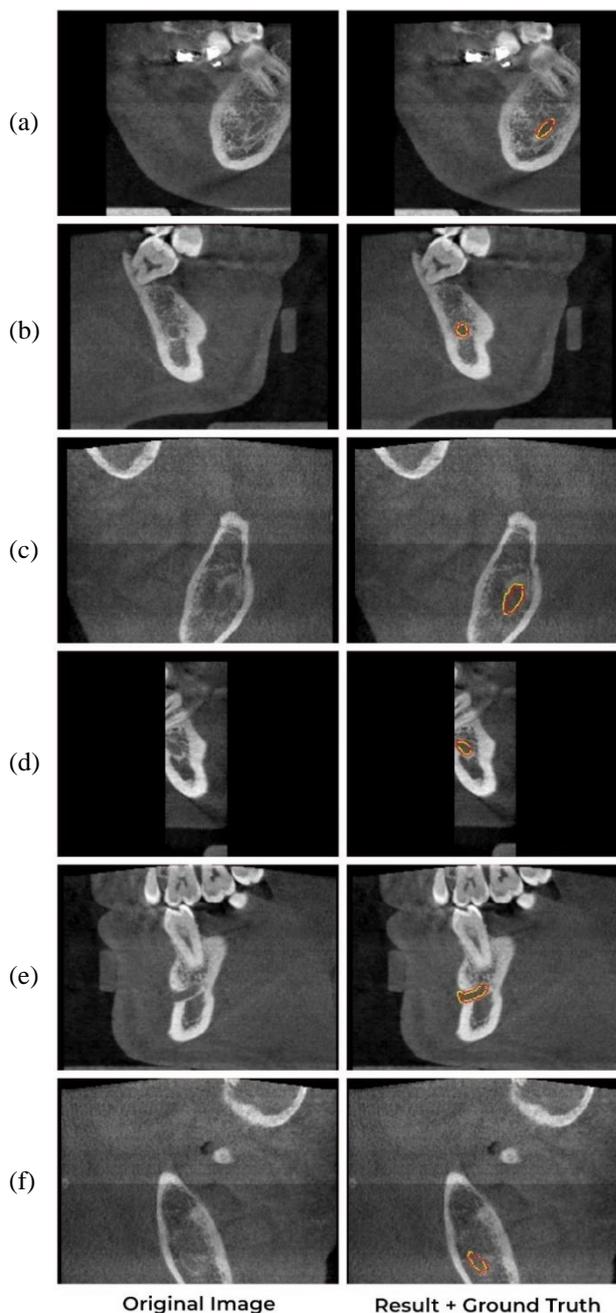


Figure. 4 Samples of mandibular canal segmentation result using the proposed RFCN + DAL

oval as shown in Fig. 4 (a), (c), (d), (e), and (f), but there is also a mandibular canal which shape like a circle as shown in Fig. 4 (b). The shape of mandibular canal quite affected the size of it. But, even if it has a similar shape, the size is various too as Fig. 4 (d) is smaller than the other oval-like mandibular canal. Fig. 4 (c)'s orientation is tilt to the top left while Fig. 4 (f) tilt to the top right and Fig. 4 (e) almost horizontal. It showed that mandibular canal has many orientations variation. The location of each mandibular canal must be different as Fig. 4 (f)'s location is at the bottom of the image and the other are almost in the middle of the image. The object

Table 2. The Architecture of each comparison method

Method	Encoder		Bridge		Decoder		Residual Building Block	Parameter
	Total Block	Filter Amount	Total Block	Filter Amount	Total Block	Filter Amount		
FCNN	3	[32, 64, 128]	1	[256]	3	[128, 64, 32]	Identity mapping	~ 2 M
ResUnet	3	[64, 128, 256]	1	[512]	3	[256, 128, 64]	Identity mapping	~ 8 M
SegNet	5	[64, 128, 256, 512, 512]	-	-	5	[512, 512, 256, 128, 64]	-	~ 24 M
U-Net	4	[64, 128, 256, 512]	1	[1024]	4	[512, 256, 128, 64]	-	~ 31 M
RFCN (ours)	3	[32, 64, 128]	1	[256]	3	[128, 64, 32]	1 × 1 convolution	~ 2 M
RFCN + DAL (ours)	3	[32, 64, 128]	1	[256]	3	[128, 64, 32]	1 × 1 convolution	~ 2 M

Table 3. Loss function and optimizer of each comparison method

Method	Loss Function		Optimizer
	Region	Boundary	
FCNN	Dice Coefficient Loss	-	Adam (LR = 0.0001)
ResUnet	Focal Loss	-	Adam (LR = 0.001)
SegNet	Weighted Cross Entropy (Median Balanced)	-	Stochastic Gradient Descent (LR = 0.1, Momentum = 0.99)
U-Net	Weighted Cross Entropy (Mean Balanced)	-	Adam (LR = 0.0001)
RFCN (ours)	Focal Loss	-	Adam (LR = 0.001)
RFCN + DAL (ours)	Focal Loss	Focal Loss	Adam (LR = 0.001)

around mandibular canal is different on each image because every patient has different oral cavity conditions as seen in Fig. 4 (e) showed more teeth compared to Fig. 4 (f). The thickness of the border around the mandibular canal is various as well as Fig. 4 (a)'s border is bolder and more visible than Fig. 4 (c). The last thing is that the grayscale intensity of mandibular canal affects its visibility as mandibular canal in Fig. 4 (a) and (b) are more visible than Fig. 4 (c) and (f). With all those different conditions, our proposed method managed to segment the mandibular canal well.

3.3.2. Method comparison

In this experiment, we compared our method with 5 deep-learning-based methods for mandibular canal segmentation. The deep-learning-based methods are our RFCN which only considered the region loss, fully convolutional neural network (FCNN), SegNet, Residual U-Net (ResU-Net), and U-Net. All of these methods, including our proposed method used U-shaped model like U-Net for the architecture. The summarized of each method's architecture were shown in Table 2.

U-Net and SegNet have a deeper network compared to other methods and only used

convolution block not residual block. Generally, convolution blocks consist of 2 convolution layers with activation function like ReLu or batch normalization after each convolution layer. While residual block, as shown in Fig. 2, is consist of convolution block and residual unit (identity mapping or 1 × 1 convolution). FCNN, ResUnet, and our RFCN used residual block in their architecture and have less parameter compared to U-Net and SegNet. The amount of parameter of each method is influence by how deep the network is and the amount of filter in each convolution layer. That's why U-Net and Segnet have much more parameters than FCNN, ResUnet, and our RFCN.

U-Net, SegNet, and ResUnet were originally used to segment other objects in another dataset. Those other datasets obviously have a different characteristic with our dataset. We tried to run those methods in their original parameter but it got underfitting when we trained U-Net and ResUnet and overfitting when we trained SegNet. So, the result was far below from our result.

Then, to optimize the capability of those methods to segment mandibular canal, we changed some of their loss function and optimizer but we still used the original architecture for all the comparison methods.

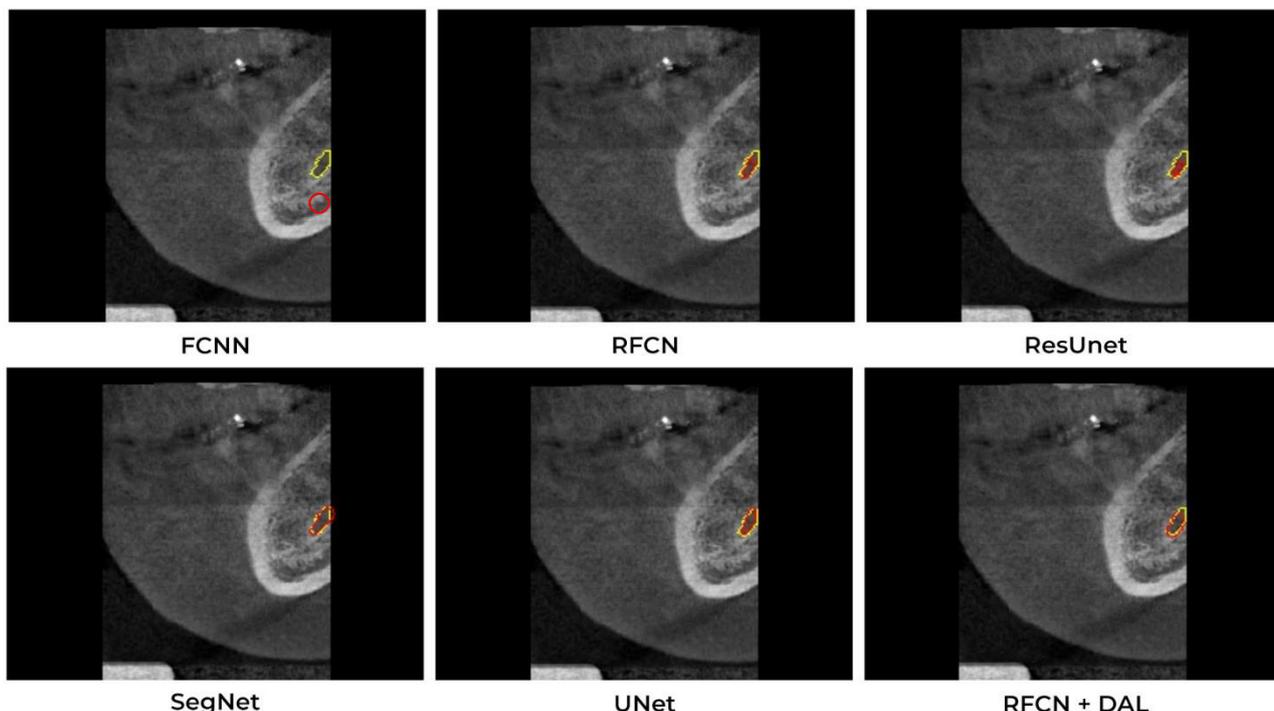


Figure. 5 Example results of the models on the first dataset. Yellow line shows the ground truth and red line shows the model’s output

Table 4. Comparison of our model with state-of-the-arts

Method	DSC	
	1st Dataset	2nd Dataset
FCNN	0.681	0.547
ResUNet	0.796	0.775
SegNet	0.814	0.716
U-Net	0.844	0.770
RFCN (ours)	0.782	0.717
RFCN + DAL (proposed method)	0.914	0.868

We summarized the loss function and optimizer we used of each method in Table 3. The method that we didn’t change its loss function and optimizer is FCNN, because it was originally used for segmenting mandibular canal in the previous research [9]. The original SegNet used VGGNet as pretrained weight, median balancing cross entropy for loss function, and stochastic gradient descent with learning rate (LR) 0.1 and momentum 0.9. We didn’t used the pretrained weight to even the other methods and changed the momentum value to 0.99. While the original U-Net used weighted pixel-wise cross entropy for loss function and stochastic gradient descent with momentum 0.99. We change the loss function to mean balancing weighted cross entropy and the optimizer to adam with learning rate 0.0001. The original ResUNet used mean squared error for the loss function and stochastic gradient descent for optimizer. We change the loss function to focal loss as mean squared error only works with balanced data [29] and

the optimizer to adam with learning rate 0.001.

We evaluated the capability of comparison method for segmenting mandibular canal based on their dice similarity coefficient score of each dataset as shown in Table 4. We used our best score in Table 1 as a comparison. We also tested the other comparison method using the model with smallest loss. The result in Table 4, showed that our model gave the highest DSC in both datasets compared with other 5 models with 0.914 on the first dataset and 0.868 on the second dataset. In the first dataset, our proposed method could exceed our RFCN without considering boundary loss, U-Net, SegNet, ResUNet, and FCNN DSC score of 0.132, 0.070, 0.100, 0.118, and 0.233 and 0.151, 0.098, 0.152, 0.093, and 0.321 in the second dataset, respectively. Our proposed method achieved a promising improvement in mandibular canal segmentation task even though our network parameters only around 1/15 of U-Net as the second-best DSC score.

The difference DSC between ResUNet and SegNet was quite thin in the first dataset while in the second dataset, ResUNet could outperform SegNet with only 1/3 parameters of SegNet. It shows that by adding residual network, the network still could segment the mandibular canal well even with less parameters. The residual network could prevent the degradation problem and help the network to extract important features with fewer number of filters.

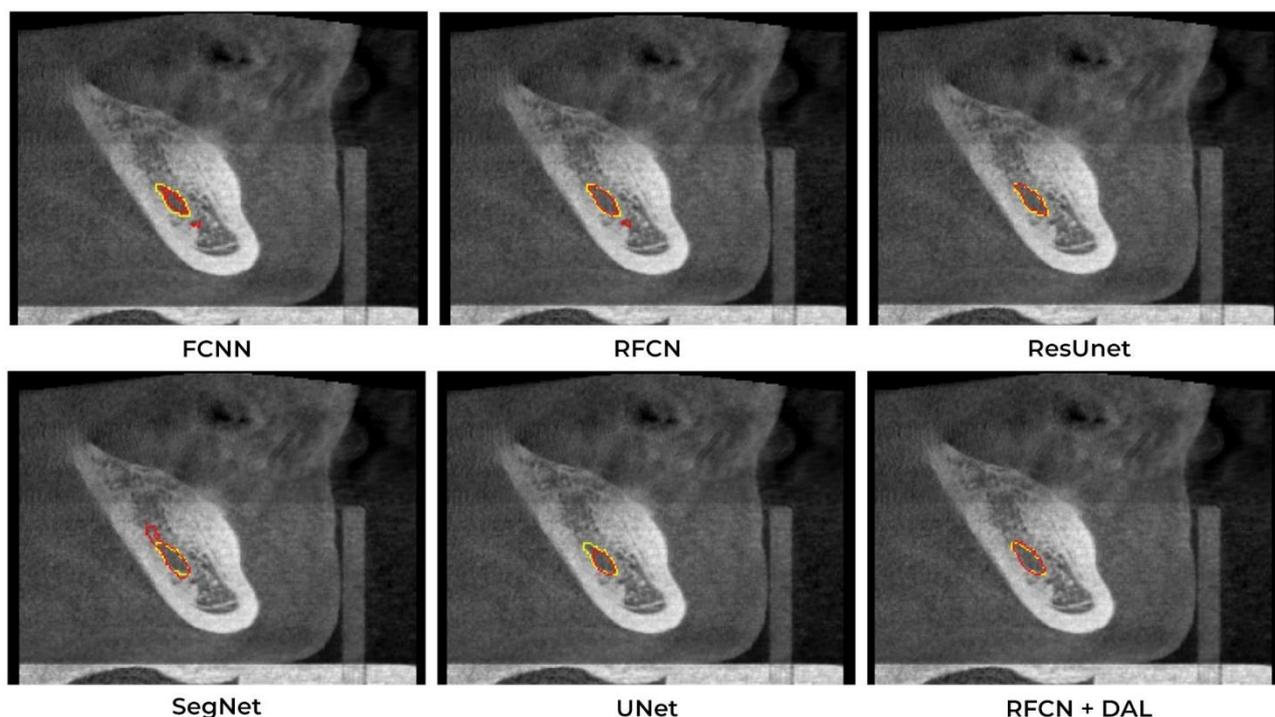


Figure. 6 Example results of the models on the second dataset. Yellow line shows the ground truth and red line shows the model's output.

Selecting the appropriate loss function was significantly affect the segmentation result. When we tried to run ResUnet with their original loss function which is mean squared error, the network run into underfitting. Then, we changed the loss function to focal loss as we used in our proposed method and it directly worked like charm. Other than focal loss, weighted cross entropy proved work well with our unbalanced data as SegNet and U-Net could still segment the mandibular canal well.

The example results from each method on the first dataset are shown in Fig. 5 and the second dataset are shown in Fig. 6 and 7. Those figures showed the position of mandibular canal on the 2D CBCT images based on the method's output and stacked it over the ground truth. The red line shows the mandibular canal based on method's output and the yellow line show the ground truth of mandibular canal.

In Fig. 5, U-Net almost successfully detect the mandibular canal while RFCN, ResUnet, and our proposed method still couldn't detect the mandibular canal on the corner of the image. FCNN could only detect it as a dot as we showed in the red circle and SegNet still detected beyond the mandibular canal's boundary. In Fig. 6, FCNN and RFCN still detected another object outside the mandibular canal and consistent as in Fig. 5, SegNet still detected far beyond the mandibular canal's boundary. U-Net only detect the part of mandibular canal while ResUnet

and our proposed method almost segmented the mandibular canal perfectly. Fig. 7 is the example of methods's result for image which didn't show the mandibular canal in the second dataset. Our proposed method and U-Net succeeded to detect that there was no mandibular canal framed, but the FCNN, RFCN, ResUnet, and SegNet still detected objects misinterpreted as mandibular canal.

Mandibular canal was a small object and looked similar with bone cavity. That could explain the reason why these models have a high chance to detected false objects. The DAL showed that by considering the boundary loss, the model performance will increase as our RFCN + DAL could exceed our RFCN's DSC score of 0.132 on the first dataset and 0.151 on the second dataset. The boundary loss could help the model to detect the pixel around the image boundaries more accurately and helped it not to detect other objects outside the mandibular canal. The boundary loss aided the model to got the better score on DSC score and segmented the object better. The used of focal loss in DAL to calculate region and boundary was proved suitable for model with unbalanced data.

4. Conclusion

The purpose of this research was to find the model to solve the problem of mandibular canal segmentation due to the highly imbalance classes between object and background. We proposed a

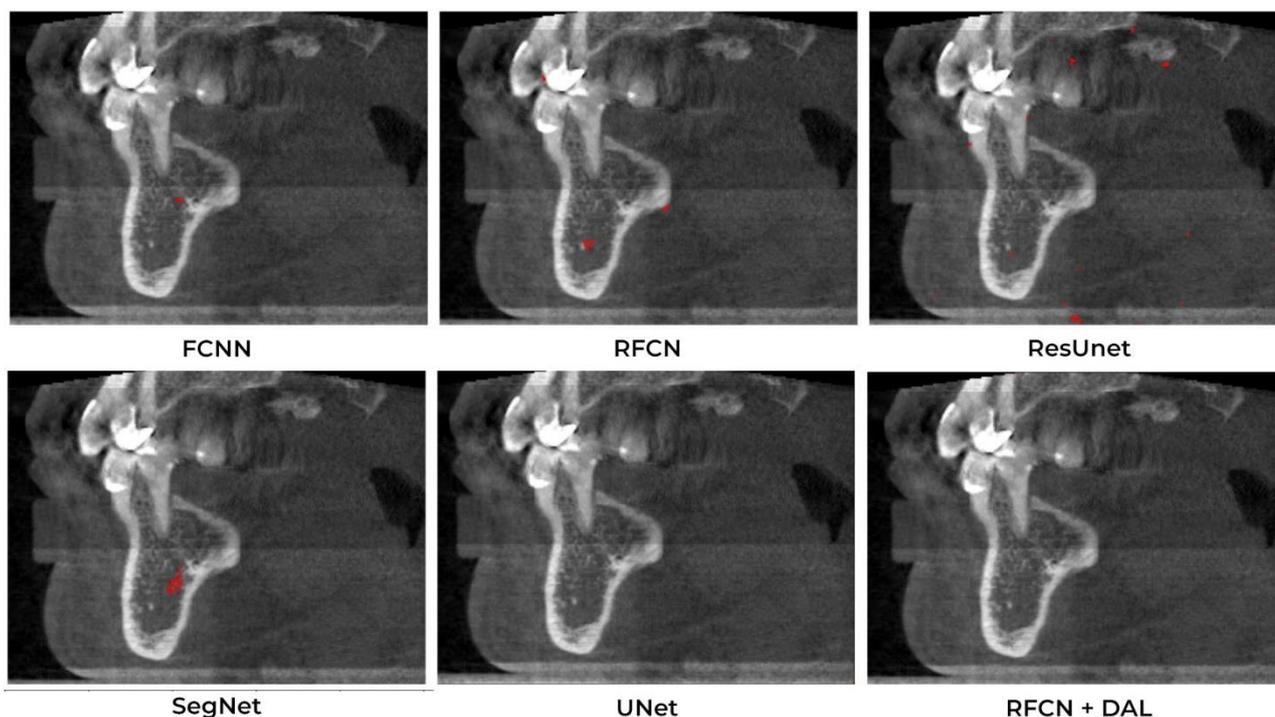


Figure. 7 Example results of the models on the second dataset which didn't show mandibular canal on the image. Red line shows the model's output

residual fully convolutional network (RFCN) with dual auxiliary loss (DAL) which considered the loss value in the region and boundary of the mandibular canal segmentation. RFCN skip connection between encoder and decoder will propagate the information and the residual unit will help the models prevent degradation information on the training phase. DAL help the RFCN detect the pixel around the image boundaries more accurately and made the model not detected others object outside the mandibular canal. The proposed RFCN + DAL model outperforms the performance of FCNN, RFCN, ResUnet, SegNet, and U-Net on both datasets. The proposed RFCN + DAL could segment the mandibular canal well even the mandibular canal has various shape, size, location, orientation, another object around the mandibular canal, thickness of the border, and the grayscale intensity. The future work of this research is finding an equation to set the weight of DAL automatically based on the pixel of the images.

Conflicts of interest

The authors declare that they have no conflict of interest in regard to this research. This article has not been published and is not under consideration for publication somewhere else.

Author contributions

Conceptualization, Yolla Faradhilla, Agus Zainal Arifin, and Eha Renwi Astuti; methodology, Yolla Faradhilla; validation, Agus Zainal Arifin and Eha Renwi Astuti; formal analysis, Yolla Faradhilla, Agus Zainal Arifin, Nanik Suciati, Rarasmaya Indraswari, and Monica Widiarsi; investigation, Yolla Faradhilla; resources, Eha Renwi Astuti and Monica Widiarsi; writing-original draft preparation, Yolla Faradhilla; writing-review and editing, Agus Zainal Arifin, Nanik Suciati, and Rarasmaya Indraswari; supervision, Agus Zainal Arifin and Nanik Suciati.

References

- [1] R. Drake, A. W. Vogl, and A. Mitchell, *Gray Dasar-Dasar Anatomi Manusia*, 1st Edition. Elsevier (Singapore) Pte. Ltd., 2014.
- [2] G. Byrne, *Fundamentals of Implant Dentistry*, 1st Editio. John Wiley & Sons, Inc, 2014.
- [3] G. Juodzbalys and M. Kubilius, "Clinical and Radiological Classification of the Jawbone Anatomy in Endosseous Dental Implant Treatment", *J. Oral Maxillofac. Res.*, Vol. 4, No. 2, pp. 1–17, 2013.
- [4] H. Mortazavi, M. Baharvand, Y. Safi, K. Dalaie, M. Behnaz, and F. Safari, "Common conditions associated with mandibular canal widening: A

- literature review”, *Imaging Sci. Dent.*, Vol. 49, No. 2, p. 87, 2019.
- [5] B. Moris, L. Claesen, S. Yi, and C. Politis, “Automated tracking of the mandibular canal in CBCT images using matching and multiple hypotheses methods”, In: *Proc. of 2012 4th Int. Conf. Commun. Electron. ICCE 2012*, pp. 327–332, 2012.
- [6] F. Abdolali, R. A. Zoroofi, and A. Biniiaz, “Fully automated detection of the mandibular canal in cone beam CT images using Lie group based statistical shape models”, In: *Proc. of 2018 25th Iran. Conf. Biomed. Eng. 2018 3rd Int. Iran. Conf. Biomed. Eng. ICBME 2018*, pp. 1–6, 2018.
- [7] F. Abdolali, R. A. Zoroofi, M. Abdolali, F. Yokota, Y. Otake, and Y. Sato, “Automatic segmentation of mandibular canal in cone beam CT images using conditional statistical shape model and fast marching”, *Int. J. Comput. Assist. Radiol. Surg.*, Vol. 12, No. 4, pp. 581–593, 2017.
- [8] D. Kainmueller, H. Lamecker, H. Seim, M. Zinser, and S. Zachow, “Automatic Extraction of Mandibular Nerve and Bone from Cone-Beam CT Data”, in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 5762 LNCS, No. Part 2, pp. 76–83, 2009.
- [9] J. Jaskari, J. Sahlsten, J. Järnstedt, H. Mehtonen, K. Karhu, O. Sundqvist, A. Hietanen, V. Varjonen, V. Mattila, and K. Kaski, “Deep Learning Method for Mandibular Canal Segmentation in Dental Cone Beam Computed Tomography Volumes”, *Sci. Rep.*, Vol. 10, No. 1, pp. 1–8, 2020.
- [10] G. H. Kwak, E. J. Kwak, J. M. Song, H. R. Park, Y. H. Jung, B. H. Cho, P. Hui, and J. J. Hwang, “Automatic mandibular canal detection using a deep convolutional neural network”, *Sci. Rep.*, Vol. 10, No. 1, p. 5711, 2020.
- [11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, In: *Proc. of Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, pp. 234–241, 2015.
- [12] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 39, No. 12, pp. 2481–2495, 2017.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, In: *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [14] Z. Zhang, Q. Liu, and Y. Wang, “Road Extraction by Deep Residual U-Net”, *IEEE Geosci. Remote Sens. Lett.*, Vol. 15, No. 5, pp. 749–753, 2018.
- [15] S. Jadon, “A survey of loss functions for semantic segmentation”, In: *Proc. of 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Oct. 2020, pp. 1–7.
- [16] S. H. Zhang, X. Dong, H. Li, R. Li, and Y. L. Yang, “PortraitNet: Real-time portrait segmentation network for mobile device”, *Comput. Graph.*, Vol. 80, pp. 104–113, 2019.
- [17] Y. D. Ma, Q. Liu, and Z. B. Quan, “Automated image segmentation using improved PCNN model based on cross-entropy”, In: *Proc. of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.*, pp. 743–746, 2004.
- [18] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection”, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 42, No. 2, pp. 318–327, 2020.
- [19] X. Xu, L. Fu, Y. Chen, R. Larsson, D. Zhang, S. Suo, J. Hua, and J. Zhao, “Breast Region Segmentation using Convolutional Neural Network in Dynamic Contrast Enhanced MRI”, In: *Proc. of 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 750–753, 2018.
- [20] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks”, *Medical Image Understanding and Analysis*, pp. 506–517, 2017.
- [21] K. Hatano, S. Murakami, H. Lu, J. K. Tan, H. Kim, and T. Aoki, “Detection of Phalange Region Based on U-Net”, In: *Proc. of 2018 18th Int. Conf. Control. Autom. Syst.*, pp. 1338–1342, 2018.
- [22] Y. Yiu, M. Aboulatta, T. Raiser, L. Opey, V. L. Flanagan, P. Z. Eulenburg, and S. A. Ahmadi, “DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning”, *J. Neurosci. Methods*, Vol. 324, p. 108307, 2019.
- [23] M. Arsalan, D. S. Kim, M. B. Lee, M. Owais, and K. R. Park, “FRED-Net: Fully residual encoder-decoder network for accurate iris segmentation”, *Expert Syst. Appl.*, Vol. 122, pp. 217–241, 2019.
- [24] G. A. Francia, C. Pedraza, M. Aceves, and S. T. Arriaga, “Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation”,

- IEEE Access*, Vol. 8, pp. 38493–38500, 2020.
- [25] H. Liu, H. Cao, E. Song, G. Ma, X. Xu, R. Jin, Y. Jin, and C. C. Hung, “A cascaded dual-pathway residual network for lung nodule segmentation in CT images”, *Phys. Medica*, Vol. 63, pp. 112–121, 2019.
- [26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *J. Mach. Learn. Res.*, Vol. 15, pp. 1929–1958, 2014.
- [27] G. S. Tran, T. P. Nghiem, V. T. Nguyen, C. M. Luong, and J. C. Burie, “Improving Accuracy of Lung Nodule Classification Using Deep Learning with Focal Loss”, *J. Healthc. Eng.*, Vol. 2019, pp. 1–9, 2019.
- [28] L. H. Shehab, O. M. Fahmy, S. M. Gasser, and M. S. E. Mahallawy, “An efficient brain tumor image segmentation based on deep residual networks (ResNets)”, *J. King Saud Univ. - Eng. Sci.*, 2020.
- [29] L. S. B. R and V. M. S, “A Weighted Mean Square Error Technique to Train Deep Belief Networks for Imbalanced Data”, *Int. J. Simul. Syst. Sci. Technol.*, Vol. 19, 2019.