DC-UNet: Rethinking the U-Net Architecture with Dual Channel Efficient CNN for Medical Images Segmentation

Ange Lou1, Shuyue Guan2, Murray Loew2

Medical Imaging and Image Analysis Laboratory
1Department of Electrical and Computer Engineering, 2Department of Biomedical Engineering
George Washington University, Washington DC, USA

Abstract

Recently, deep learning has become much more popular in computer vision applications. The Convolutional Neural Network (CNN) has brought a breakthrough in image segmentation, especially for medical images. In this regard, the U-Net is the predominant approach to the medical image segmentation task. The U-Net not only performs well in segmenting multimodal medical images generally, but also in some difficult cases. We found, however, that the classical U-Net architecture has limitations in several respects. Therefore, we applied modifications: 1) designed efficient CNN architecture to replace encoder and decoder, 2) applied residual module to replace skip connection between encoder and decoder to improve, based on the state-of-the-art U-Net model. Following these modifications, we designed a novel architecture -- DC-UNet, as a potential successor to the U-Net architecture. We created a new effective CNN architecture and built the DC-UNet based on this CNN. We have evaluated our model on three datasets with difficult cases and have obtained a relative improvement in performance of 2.90%, 1.49%, and 11.42% respectively compared with classical U-Net. In addition, we used the Tanimoto similarity measure to replace the Jaccard measure for gray-to-gray image comparisons.

Keywords: convolutional neural network, MultiResUnet, deep learning, medical image segmentation, computer aided diagnosis, DC-UNet

1. INTRODUCTION

The goals of medical image analysis are to provide efficient diagnostic and treatment processes for radiologists and clinicians [1]. Medical imaging devices such as X-ray, CT, and MRI can provide noninvasive information about disease, abnormality, and anatomy in the human body. The large amount of data (and often artifact) in medical images requires that we process images to extract effective information [2] from them. Medical image processing has contributed substantially in medical applications; for example, image segmentation, image registration, and image-guided surgery are widely used in medical treatment.

An essential technique in medical image processing is image segmentation, which, among other applications, is used to identify the region of interest (ROI) through some automatic and semi-automatic methods [3]. There are many traditional algorithms designed to segment tissues or body organs. These methods can be classified as: region-based, edge-based, threshold- and feature-based clustering [4]. Region-based algorithms find a group of connected pixels that have similar properties. Segmentation based on edge detection refers to the boundaries where there is an abrupt change in the intensity or brightness value of the image. There are various edge detection algorithms like Sobel detector [5], Canny detector [6] and Fuzzy Inference System (FIS) [7]. Generally, before applying edge detection, we need to pre-process and enhance the images. Thresholding is a simple and powerful technique for segmenting images having large contrast between objects and background. The thresholding operation converts a multi-level gray image into a binary image through an appropriate threshold T and divides image pixels into several regions and separates objects from background. Otsu’s [8] is the most popular thresholding method. Clustering is the process to group objects based on some similar properties so that each cluster contains similar objects. Examples of clustering methods include K-means [9] and Fuzzy C Means (FCM) [10].

Those traditional segmentation methods, however, do not work well in difficult cases. In the case of the CVC-ClinicDB [24] endoscopy dataset, for example, the shape, size, and boundary of polyps are totally different, and some polyps with vague boundaries cannot be detected by traditional segmentation approaches. Recently, deep learning methods have been shown to outperform previous state-of-the-art machine learning techniques for several computer vision tasks [11]. According to the main principle of recent deep learning segmentation methods, they contain two categories: region-based
semantic segmentation and full-convolutional-network (FCN)-based semantic segmentation [12]. For region-based deep learning methods, Region with CNN feature (RCNN) [13] is a well-known model. RCNN utilizes features from a CNN detector to address those complicated tasks. Moreover, any CNN structure can be used as a detector in RCNN model, such as AlexNet [14], VGG [15], GoogLeNet [16] and ResNet [17]. The FCN-based [18] is a pixel-wise segmentation method. Compared with region-based methods, FCN-based does not need to extract the proposed regions. The FCN is an extension of the classical CNN and uses a decoder-like part to generate the segmentation mask. After FCN, large numbers of encoder-decoder models have been applied to segmentation, such as SegNet [19] and U-Net [20]. The U-Net and U-Net-like models have been successfully used to segment various biomedical images, such as liver [21], skin lesions [22], and vessels [23].

In this paper, we develop a novel model called Dual Channel U-Net (DC-UNet); it is an enhanced version of U-Net, which is the most popular and successful deep learning model for biomedical image segmentation to date. We believe the DC-UNet will significantly contribute to medical image segmentation. We tested the DC-UNet by using a variety of medical images. Our results show that DC-UNet surpasses the classical U-Net model in all cases, even if DC-UNet has slightly fewer parameters.

2. THE DUAL-CHANNEL U-NET

The U-Net [20] has been a remarkable and the most popular architecture in medical image segmentation and the MultiResUNet [30] can provide a much better output than the U-Net, because it can provide different-scale features. For some extremely challenging medical image cases, however, the MultiResUNet cannot perform well, such as for fuzzy objects and in the presence of interference of backgrounds (e.g., visibility of medical equipment). The goal of the MultiRes block, as shown in Figure. 1(a), is to provide different-scale features to help separate objects from the whole image.

Our previous work (using MultiResUNet to segment the breast region from infrared (IR) images) has shown that some small-breast IR images do not have clear breast boundaries and some segmentation results will be influenced by other

![Diagram of MultiRes Block](image1)

![Diagram of Dual-Channel Block](image2)

![Diagram of Res-Path](image3)

Fig. 1. (a) MultiRes Block. (b) Dual-Channel block (c) Res-Path

Our previous work (using MultiResUNet to segment the breast region from infrared (IR) images) has shown that some small-breast IR images do not have clear breast boundaries and some segmentation results will be influenced by other
interferences, such as patient’s belly and parts of medical equipment. Those factors influenced the segmentation results from MultiResUNet. We solved this problem by designing a more effective CNN architecture to extract more spatial features. To compare the segmentation results from the classical U-Net and MultiResUNet, we can easily find that different-scale features greatly help segmentation. Thus, we assumed that those most challenging tasks would be solved if we can provide more different-scale (more effective) features.

![Fig. 2. Architecture of DC-UNet](image)

**Table 1. Details of DC-UNet**

| DC-UNet | Block | Layer (left) | #Filters | Layer (right) | #Filters |
|---------|-------|-------------|----------|---------------|----------|
| DC Block 1 | Conv2D(3,3) | 8 | Conv2D(3,3) | 8 |
| DC Block 2 | Conv2D(3,3) | 17 | Conv2D(3,3) | 17 |
| DC Block 3 | Conv2D(3,3) | 53 | Conv2D(3,3) | 53 |
| DC Block 4 | Conv2D(3,3) | 71 | Conv2D(3,3) | 71 |
| DC Block 5 | Conv2D(3,3) | 142 | Conv2D(3,3) | 142 |
| Res Path 1 | Conv2D(3,3) | 32 | Res Path 2 | Conv2D(3,3) | 64 |
| Res Path 3 | Conv2D(3,3) | 128 |
| Res Path 4 | Conv2D(3,3) | 256 |

Based on this assumption, we noticed there was a simple residual connection in the MultiRes block. As noted in that work, the residual connection here provides only a few additional spatial features, which may be not enough for some most-challenging tasks. Different-scale features have already shown potential in medical image segmentation. Meanwhile, MultiResUNet proves that Res-path provides only small improvement of performance compared with MultiRes block. Thus, to overcome the problem of insufficient spatial features, we took a sequence of three $3 \times 3$ convolutional layers to replace the residual connection in MultiRes block. We called this Dual-Channel block, as shown in Figure. 1(b). We applied the same connection (Res-Path, Figure. 1(c)) between encoder and decoder as in the MultiResUNet.
utilized the Res-Path and Dual-Channel block to build a new U-Net architecture -- DC-UNet -- whose architecture is illustrated in Figure 2.

Each channel in Dual-Channel block has half the filter numbers of MultiRes block: \{32, 64, 128, 256, 512\}. \(W\) is each layer’s filter number. And \(W\) meets the equation (1) as mentioned before. We applied the same \(U\) and \(\alpha\) value, and the filter number of three \(3 \times 3\) are also divided into \(\frac{W}{6}\), \(\frac{W}{3}\), and \(\frac{W}{2}\). Moreover, the number of filters in Res-path are \{32, 64, 128, 256\}. And the number of each layer’s filter is shown in Table 1. All convolutional layers in the DC-UNet are activated by the ReLU function and batch normalization is used to avoid overfitting. The final output layer is activated by Sigmoid function.

3. EXPERIMENTS

In the experiments, the network models were built by using Keras [32] with Tensorflow backend [33] in Python 3 [31]. The experiments were conducted on a desktop computer with Intel core i7-9700K processor (3.6 GHz) CPU, 16.0 GB RAM, and NVIDIA GeForce RTX 2070 GPU.

3.1 Baseline model

In these experiments, we chose the U-Net as the baseline model and compared its performance with MultiResUNet and DC-UNet. To show the advantage in parameters, we implemented the classical U-Net with five-stage encoder and decoder, and the filter numbers are \{64, 128, 256, 512, 1024\}. For the MultiResUNet and DC-UNet, we also used five-stage encoder and decoder. And each layer’s filter number of the DC-UNet can be found in Table 1.

3.2 Pre-processing

The goal of our experiments is to test the performance of the DC-UNet and compare it with the classical U-Net and MultiResUNet. The pre-processing we applied for the IR thermography database converted 16-bit images to 8-bit and resized the image to 256 x 128. Due to the limitation of GPU memory, the pre-processing for other databases is to resize the images to no larger than 256 x 256.

3.3 Training

The goal of semantic segmentation is to predict whether a pixel belongs to the object. Therefore, this problem can be considered as a pixel-wise binary classification problem. Hence, we chose the binary cross-entropy as loss function and minimized it. For the input image \(X\), the prediction of model is \(\hat{y}\) and the ground truth is \(y\). Thus, the binary cross-entropy is defined as:

\[
\text{Cross Entropy}(y, \hat{y}) = \sum_{x \in X} - (y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))
\]  

(1)

For a batch containing \(n\) images, the loss function \(J\) is defined as:

\[
J = \frac{1}{n} \sum_{i=1}^{n} \text{Cross Entropy}(y, \hat{y})
\]  

(2)

We trained those models using the Adam optimizer [34] with the parameter \(\beta_1 = 0.9\) and \(\beta_2 = 0.999\). Epochs are varied by datasets.

3.4 Measurement metric

To evaluate the performance of segmentation, we need a method to compare the segmented region with the ground truth region. Since the final layer is activated by a Sigmoid function, it produces an output in the range \([0, 1]\). Therefore, we cannot compare the output with ground truth directly because the ground truth data are binary images. Usually, image
thresholding from grayscale to binary (binarization) [35] will lose much information. Using our previous infrared breast region segmentation study for example, after converting 16-bit images to 8-bit (pixel value range in [0, 255]), there are many usable comparison methods for two images:

- Binary vs Binary: Jaccard Similarity (JS) [36]
- Gray vs Gray: Mean Absolute Error (MAE) [37]; Tanimoto Similarity [38] (Extended Jaccard Similarity); Structural similarity (SSIM) [39]

The Tanimoto similarity, also called extended JS, can be considered as a grayscale version JS. For a binary image, JS compares images by union and intersection operations. The union operation could be considered as sum of products. For two set A and B:

\[ |A \cap B| = \sum a_i b_i \]  

(3)

Where \( a_i \in A, b_i \in B \). This equation holds if \( a_i, b_i \in \{0, 1\} \), which are binary values. But if \( a_i, b_i \) are not binary, we use the sum of products (right part) instead of the union operation. Since:

\[ |A \cap A| = \sum a_i^2 \]  

(4)

And,

\[ |A \cup B| = |A| + |B| - |A \cap B| = \sum (a_i^2 + b_i^2 - a_i b_i) \]  

(5)

For gray-gray comparison, according to \( JS(A, B) \), the value of Tanimoto similarity is:

\[ T(A, B) = \frac{\sum a_i b_i}{\sum (a_i^2 + b_i^2 - a_i b_i)} \]  

(6)

By definition, Tanimoto similarity is similar to JS but more general than JS and has wider applications. Therefore, it is a good alternative method for segmentation evaluation.

3.5 Cross validation

Cross-Validation is widely used to test a model’s performance. In the k-fold cross-validation test, the dataset \( D \) is randomly split into \( k \) mutually exclusive subsets \( D_1, D_2, \ldots, D_k \) of approximately equal size [40]. The model is run \( k \) times; for each time, one of the \( k \) subsets is chosen as the validation set and all others as the training set. We estimated the performance of the model via overall results from \( k \) times training.

4. DATASETS

Compared with traditional computer vision datasets, current medical imaging datasets are more challenging. Expensive medical equipment, complex image acquisition pipelines, diagnosis by experts and tedious manual labeling – they all make medical datasets difficult to build. Currently, there are some public medical imaging benchmark datasets containing medical images and their ground truth. We have selected two public datasets and our own infrared breast dataset to test the performance of the three U-Net-based models. The datasets used in the experiments are briefly described in Table 2.

| Modality                  | Dataset            | No. of images | Original resolution | Input resolution |
|---------------------------|--------------------|---------------|---------------------|------------------|
| Thermography              | Our IR breast      | 450           | Variable            | 256 x 128        |
| Electron microscopy       | ISBI-2012          | 30            | 512 x 512           | 256 x 256        |
| Endoscopy                 | CVC_ClinicDB       | 612           | 384 x 288           | 128 x 96         |

4.1 Infrared breast images

We collected infrared images using the N2 Imager (N2 Imaging System, Irvine, Calif.). Patients diagnosed with breast
cancer and healthy volunteers are imaged by the infrared camera for 15 minutes to observe the cooling of the breast tissue. This dynamic thermography monitors the temporal behavior of breast thermal patterns, which in our case is the cooling of breast tissue over time. The patients and volunteers keep sitting with both arms raised on supports, with the camera positioned approximately 25 inches away from the breasts (frontal view). The imager starts to capture images immediately after the patient undressed; images were taken every minute as the breast surface cooled.

Our breast dataset contains 450 infrared images from 14 patients and 16 healthy volunteers; all images contain background objects and noise. Each participant was imaged for a total time of 15 minutes, capturing one image every minute (15 images per participant). The original resolution of the images ranges from $540 \times 260$ to $610 \times 290$; we have resized them to $256 \times 128$ due to limitation of memory.

4.2 Electron microscopy (EM)

To show the performance of the new model in electron microscopy (EM) images, we choose the dataset of the ISBI-2012 challenge: 2D EM segmentation [41]. This dataset contains 30 images in its training set from a serial section Transmission Electron Microscopy (ssTEM) of the Drosophila first instar larva ventral nerve cord [42]. Due to the testing set does not contain ground truth, we totally chose 30 images in training set as dataset. The resolution of images is $512 \times 512$, we have resized the images to $256 \times 256$ due to the limitation of memory. And for EM segmentation experiment, we took 5-fold cross-validation.

4.3 Endoscopy images

To show the performance of new model in endoscopy images, we chose CVC-ClinicDB [43] as one dataset. These images were extracted from the colonoscopy videos. This dataset contains total 612 images with ground truth, and their original resolution was $384 \times 288$. We resized the images to $128 \times 96$ for training due to the limitation of memory.

5. RESULTS

5.1 Models’ parameters

To evaluate the performance of DC-UNet, we first compared the performance of U-Net with different parameters on infrared breast datasets. Then, we designed experiments to compare performance of U-Net, MultiResUNet, and DC-UNet on three medical datasets; the numbers of parameters appear in Table 3.

| Model       | Parameters   |
|-------------|--------------|
| U-Net*      | 7750821      |
| U-Net       | 31,031,685   |
| MultiResUNet| 29,061,741   |
| DC-UNet     | 10,069,640   |

5.2 Results for infrared breast images

The infrared breast dataset contains 450 images and was divided into 30 subsets ($k = 30$) by participant. Every model has been trained 50 epochs at each run. The overall average accuracies of U-Net*, U-Net, MultiResUNet and DC-UNet are 88.70%, 89.80%, 91.47% and 92.71%, respectively, after applying the 30-Fold cross validation for three models. The average accuracies and standard deviations for each participant are shown in Table 5, the DC-UNet performs better than the other models for most test cases. Table 4 shows the specific average accuracy values of each patients.

From the results in Table 4 and Table 5, DC-UNet provides more accurate segmentation results both for simple and challenging cases. For example, to a simple case V7, the segmentation accuracies of U-Net, MultiResUNet and DC-UNet were 92.47%, 93.86% and 95.38%, respectively. DC-UNet gives the best segmentation results compared to the other models because of using our new DC blocks.
Table 4. Average segmentation accuracy for each sample. Bold value is the maximum for each participant.

|     | P1  | P2  | P3  | P4  | P5  | P6  | P7  | P8  | P9  | P10 | P11 | P12 | P13 | P14 | V1  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| U-Net* | 84.5 | 86.4 | 86.8 | 92.7 | 88.2 | 89.8 | 88.9 | 79.1 | 87.6 | 93.5 | 84.8 | 89.5 | 92.5 | 86.1 | 92.3 |
| U-Net | 85.2 | 87.9 | 87.0 | 94.4 | 89.1 | 91.1 | 89.7 | 79.8 | 89.4 | 94.0 | 87.0 | 90.1 | 93.3 | 87.7 | 93.1 |
| MultiResUNet | 88.2 | 92.9 | 90.5 | 96.7 | 91.8 | 91.6 | 89.0 | 80.0 | 90.5 | 95.3 | 84.0 | 95.1 | 93.2 | 87.0 | 95.9 |
| DC-UNet | 90.3 | 94.7 | 91.1 | 96.7 | 90.7 | 92.4 | 90.0 | 83.4 | 91.4 | 95.5 | 92.6 | 96.2 | 95.9 | 87.6 | 95.7 |

V2  | V3  | V4  | V5  | V6  | V7  | V8  | V9  | V10 | V11 | V12 | V13 | V14 | V15 | V16 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| U-Net* | 91.4 | 89.5 | 91.0 | 87.8 | 84.1 | 91.7 | 82.6 | 93.8 | 82.5 | 92.7 | 93.3 | 88.2 | 90.4 | 92.3 | 87.2 |
| U-Net | 92.2 | 90.4 | 92.7 | 88.7 | 85.5 | 92.5 | 84.3 | 95.1 | 82.7 | 94.1 | 94.8 | 89.8 | 91.8 | 93.0 | 87.9 |
| MultiResUNet | 93.0 | 90.8 | 92.6 | 90.3 | 90.1 | 93.9 | 88.2 | 96.0 | 88.4 | 96.5 | 95.7 | 89.6 | 93.2 | 95.2 | 89.0 |
| DC-UNet | 93.8 | 93.6 | 94.1 | 91.4 | 91.1 | 95.4 | 88.4 | 96.9 | 89.3 | 97.1 | 95.5 | 90.3 | 94.1 | 95.5 | 91.0 |

Table 5. Average segmentation accuracy and standard deviation for totally 30 samples

| Model       | Average accuracy | Standard deviation |
|-------------|------------------|--------------------|
| U-Net*      | 88.7%            | 0.0256             |
| U-Net       | 89.8%            | 0.0241             |
| MultiResUNet | 91.5%          | 0.0237             |
| DC-UNet     | 92.7%            | 0.0215             |

For the challenging cases, DC-UNet also gives inspiring results. For the example of P11, its image contains many interferences like medical equipment and other parts of the body as shown in Figure. 3. The segmentation accuracies of U-Net, MultiResUNet and DC-UNet are 86.47%, 84.01% and 92.62%, respectively. We can find that only DC-UNet can clearly separate breast boundary from belly. Thus, the DC-UNet gives an outstanding result than other models.

![Segmentation result of patient 11. (a) Original image (b) Manual ground-truth (c) U-Net (d) MultiResUNet (e) DC-UNet](image)

5.3 Results for microscopy images

For the electron microscopy (EM) dataset, we have performed 5-fold cross-validation and compared the performance of DC-UNet with MultiResUnet and the baseline U-Net. Every model has been trained for 50 epochs for each run and recorded the Tanimoto accuracy. The results of EM dataset were shown in Table 6. From the Table 6, we can find that the DC-UNet gives the best results for all cases.
Table 6. The result of EM through a 5-fold cross-validation. Bold values are the maximum for each case.

| Model     | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Average | Std   |
|-----------|--------|--------|--------|--------|--------|---------|-------|
| U-Net     | 90.75  | 89.31  | 91.13  | 91.60  | 91.13  | 91.13   | 0.0157|
| MultiResUNet | 90.83  | 91.16  | 92.64  | 93.11  | 92.08  | 91.96   | 0.0121|
| DC-UNet   | 91.79  | 91.27  | 93.21  | 93.44  | 93.38  | 92.62   | 0.0092|

5.4 Results for endoscopy images

For the endoscopy dataset, we performed 5-fold cross-validation and compared the performance of DC-UNet with MultiResUNet and the baseline U-Net. In the experiments, each model has been trained for 150 epochs for each run and the Tanimoto accuracy computed. The results for the endoscopy dataset were shown in Table 7.

Table 7. The result of endoscopy through a 5-fold cross-validation. Bold values are the maximum for each case.

| Model     | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Average | Std   |
|-----------|--------|--------|--------|--------|--------|---------|-------|
| U-Net     | 74.03  | 70.81  | 67.96  | 63.26  | 71.52  | 69.52   | 0.2950|
| MultiResUNet | 81.82  | 80.34  | 79.57  | 74.23  | 78.66  | 78.92   | 0.2738|
| DC-UNet   | 83.11  | 82.51  | 81.10  | 78.14  | 79.84  | 80.94   | 0.2493|

Fig. 4. Segment images with small objects. (a) Original image (b) Ground truth (c) U-Net (0%) (d) MultiResUNet (11.76%) (e) DC-UNet (69.00%)

From Table 7, we can find that MultiResUNet gives much better results than U-Net in this challenging dataset. The average accuracy has been improved 9.4%. The DC-UNet can successfully segment images with vague boundaries and successfully detect small objects in images, as shown in Figure 4. The segmentation accuracy has been improved 11.42% compared to the U-Net. For some easy cases in CVC-ClinicDB, results of MultiResUNet and DC-UNet are similar, but much better than classical U-Net.
6. DISCUSSION

6.1 Comparison of three models

From these experiments, DC-UNet shows great potential in multimodal medical image segmentation. In our infrared breast dataset, DC-UNet can separate the belly and breast region even though they have similar temperatures, and DC-UNet also provides a more accurate contour. From the results of EM dataset, the DC-UNet shows good robustness to noise. In the CVC-ClinicDB dataset, DC-UNet shows a great ability, which MultiResUnet and U-Net do not have, to segment small objects and vague boundaries without any data augmentation techniques.

In addition, DC-UNet is more efficient because there are far fewer parameters than in MultiResUNet and the classical U-Net. The number of parameters is related to convolutional kernel size and numbers of channels of input and output. In our DC block, each channel’s number of filters is half those of the corresponding MultiRes block. After passing the add layer, we calculate sum of these two channels instead of concatenating. Thus, the dimension of the current output layer and next input layer are half of the corresponding layer in MultiRes block. Moreover, half output dimension in DC block also leads to the number of filters of Res-Paths being half of that in MultiResUNet. Based on reduced dimension of input and output, the number of parameters in DC-UNet are much fewer than MultiResUNet and U-Net. Nevertheless, it contains double the multi-resolution features that make the results better than those of the compared models.

6.2 Improvements and future work

To get better results, data augmentation [44] like flipping, rotation, and randomly cropping to enlarge the datasets, and image enhancement [24] are very helpful techniques. Data augmentation operations can help models avoid overfitting during training [25]. Moreover, objects in medical images sometimes do not have clear boundaries because of poor illumination, noise, and/or tissue properties. Thus, the histogram equalization technique, which can improve the contrast, such as CLAHE [26] would be greatly helpful.

Despite data augmentation and image enhancement techniques, there are also potentials in the dual-channel CNN architectures. In our experiments, we use only a dual-channel model for segmentation. Adding more channels like blocks in ResNeXt [27] will provide more effective features, but it will cause the increment of parameters and floating-point operations (FLOPs). Moreover, there are also other versions’ Inception module, such as Inception-v4 [29] and Inception-v3 [28], which proposed using asymmetric convolution to replace the original convolution kernel. For example, the $3 \times 3$ convolution operator can be replaced by a $3 \times 1$ convolution following a $1 \times 3$ convolution to minimize the parameters further. In the future, we will test our model on more datasets. Moreover, we will study how data augmentation and preprocessing could improve the model’s performance.

7. CONCLUSION

In this work, we analyzed the classical U-Net and the recent MultiResUNet architecture and found potential improvements. We note that the results of our own infrared breast dataset still have many limitations when using classical U-Net and MultiResUNet. The author of MultiResUNet paper [30] has verified that Res-Path can slightly improve the segmentation accuracy. Thus, we designed the Dual-Channel CNN block to give more effective features with fewer parameters to overcome those limitations. To incorporate this dual-channel CNN architecture with Res-Path, we developed a novel U-Net-like architecture -- DC-UNet.

We selected two public medical datasets and our own infrared breast dataset to test and compare the performance of these three models. Each dataset contains some challenging cases. The infrared breast dataset contains small-size breast images with unclear boundaries. Some images in the ISBI-2012 Electron Microscopy dataset contain many interferences like noise and other parts of cells that will influence the model in recognizing the boundaries. For colon endoscopy images in CVC-ClinicDB, the boundaries of polyps are very vague and difficult to distinguish and the shapes, sizes, structures, and positions of polyps are quite variable. Those factors make this dataset most challenging.

For those challenging cases, the performance of DC-UNet was better than of MultiResUNet and U-Net. For the infrared breast, ISBI-2012, and CVC-ClinicDB datasets, a relative improvement of segmentation accuracy 2.90%, 1.49%, and
11.42%, respectively, has been observed in using DC-UNet compared to U-Net. And DC-UNet also has 1.20%, 0.66%, and 2.02% improvements, respectively, compared to Multi-ResUNet. Besides higher segmentation accuracies DC-UNet achieved, the segmentation results are much closer to the ground truth by observation. In some cases, U-Net and MultiResUNet tend to under-segment and even miss the objects completely. In contrast, the DC-UNet is more reliable and robust. DC-UNet can detect vague boundaries and avoid the interference of noise. Even for the challenging cases, the DC-UNet shows a stronger ability to capture fine details. Therefore, we believe that the DC-UNet architecture can be an effective model for medical image segmentation.

REFERENCE

[1] Gao, J., Jiang, Q., Zhou, B., & Chen, D. (2019). Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview. Mathematical Biosciences and Engineering, 16(6), 6536.
[2] Song, H., Nguyen, A. D., Gong, M., & Lee, S. (2016). A review of computer vision methods for purpose on computer-aided diagnosis. J. Int. Soc. Simul. Surg, 3, 1-8.
[3] Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rad, A. E., Rehman, A., & Uddin, M. (2014). Medical image segmentation methods, algorithms, and applications. IETE Technical Review, 31(3), 199-213.
[4] Kuruvilla, J., Sukumaran, D., Sankar, A., & Joy, S. P. (2016, March). A review on image processing and image segmentation. In 2016 international conference on data mining and advanced computing (SAPIENCE) (pp. 198-203). IEEE.
[5] Gupta, S., & Mazumdar, S. G. (2013). Sobel edge detection algorithm. International journal of computer science and management Research, 2(2), 1578-1583.
[6] Ding, L., & Goshthesby, A. (2001). On the Canny edge detector. Pattern Recognition, 34(3), 721-725.
[7] Kikuchi, S., & Chakroborty, P. (1992). Car-following model based on fuzzy inference system. Transportation Research Record, 82-82.
[8] Vala, H. J., & Baxi, A. (2013). A review on Otsu image segmentation algorithm. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(2), 387-389.
[9] Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.
[10] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3), 191-203.
[11] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018.
[12] Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. International journal of multimedia information retrieval, 7(2), 87-93.
[13] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
[14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[16] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[18] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
[19] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.
[20] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[21] Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., ... & Sommer, W. H. (2016, October). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D
conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 415-423). Springer, Cham.

[22] Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., ... & Halpern, A. (2018, April). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168-172). IEEE.

[23] Xiao, X., Lian, S., Luo, Z., & Li, S. (2018, October). Weighted Res-UNet for High-Quality Retina Vessel Segmentation. In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 327-331). IEEE.

[24] Maini, R., & Aggarwal, H. (2010). A comprehensive review of image enhancement techniques. *arXiv preprint arXiv:1003.4053*.

[25] Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2016, November). Understanding data augmentation for classification: when to warp? In *2016 international conference on digital image computing: techniques and applications (DICTA)* (pp. 1-6). IEEE.

[26] Reza, A. M. (2004). 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*, 38(1), 35-44.

[27] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).

[28] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[29] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[30] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74-87.

[31] Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, 9(3), 10-20.

[32] Chollet, F. (2015). Keras.

[33] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16) (pp. 265-283).

[34] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[35] Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1), 146-166.

[36] Kosub, S. (2019). A note on the triangle inequality for the jaccard distance. *Pattern Recognition Letters*, 120, 36-38.

[37] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?--Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3), 1247-1250.

[38] Rogers, D. J., & Tanimoto, T. T. (1960). A computer program for classifying plants. *Science*, 132(3434), 1115-1118.

[39] Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003, November). Multiscale structural similarity for image quality assessment. In *the Thiry-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003 (Vol. 2, pp. 1398-1402). Ieee.

[40] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*(Vol. 14, No. 2, pp. 1137-1145).

[41] Arganda-Carreras, I., Turaga, S. C., Berger, D. R., Cireşan, D., Giusti, A., Gambardella, L. M., ... & Liu, T. (2015). Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in neuroanatomy*, 9, 142.

[42] Cardona, A., Saaßfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., ... & Hartenstein, V. (2010). An integrated micro-and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. *PLoS biology*, 8(10), e1000502.

[43] Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., & Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43, 99-111.

[44] Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.