International Journal of Artificial Intelligence
p-ISSN: 2407-7275, e-ISSN: 2686-3251

Original Research Paper

An Innovative Technique for Medical Image Segmentation Using Convolutional Neural Networks Optimized Through Stochastic Gradient Descent

Mohammad Taheri[1*], Faezeh Sadeghi[2], Abbas Koochari[2]

[1] *Department of Computer Engineering, Faculty of Engineering, Damghan University. Damghan, Iran.*

[2] *Department of Computer Engineering, Faculty of Engineering, Islamic Azad University, Science and Research Branch. Tehran, Iran.*

Article History
Received:
13.05.2024

Revised:
03.06.2024

Accepted:
20.06.2024

***Corresponding Author:**
Mohammad Taheri
Email:
mohammad-taheri@du.ac.ir

Abstract: Medical image segmentation is crucial due to its essential role in disease therapy. Various challenges such as hair artifacts, illumination variations, and different imaging acquisitions complicate this task. In this paper, we introduce a novel convolutional neural network (CNN) architecture designed to address these challenges. We also compare our method with two well-known architectures, Unet and FCN, to demonstrate the superiority of our approach. Our results, evaluated using four metrics, accuracy, dice coefficient, Jaccard index, and sensitivity show that our method outperforms the other two. We employed Jaccard distance and binary cross-entropy as the loss functions and used SGD+Nesterov as the optimization algorithm, which proved more effective than the Adam optimizer. In the preprocessing step, we included image resizing to speed up the process and image augmentation to enhance the evaluation metrics. As a postprocessing step, we applied a threshold technique to the algorithm's outputs to increase the contrast of the final images. This method was tested on two well-known and publicly available medical image datasets: PH2 for melanoma detection and Chest X-ray images for detecting chest lesions or COVID-19.

Keywords: Convolutional Neural Networks, Covid-19, Deep Learning, Medical Image Segmentation, Melanoma.

1. Introduction

Medical image analysis is the first step for detecting different kinds of diseases by the physicians. Medical image segmentation is one the most important and vital tasks toward medical image analysis by separating the tumors or lesions from the other healthy textures. Covid19 is an infectious disease that contaminates lungs. because of this, physicians need an automatic method that can segment the lungs in Xray images to detect better the infectious parts of lungs.

In this paper, we have rendered a new architecture of convolutional neural networks toward medical image segmentation which has been implemented on the two practical datasets for skin lesion and chest segmentation. Additionally, we have implemented FCN and U-net architectures on these datasets and compared the results by the outputs of our method to evaluate its robustness. Finally, we discovered that our method could outperformed these two famous methods for medical image segmentation. Meanwhile, there are some features such as irregular and fuzzy borders, low contrast between the lesions and other surrounding textures, hair artifacts, blood vessels in medical images that make medical image segmentation more challenging.

In this paper, firstly, we discuss about the previous works in section II, then we talk about all of the details of our proposed method for medical image segmentation in section III, then we report all of the experimental results and design in section IV. Finally, we display the results in section V and conclude our study in section VI.

2. Literature Review

In the past years, researches developed many algorithms to solve medical image segmentation task. Some of these efforts include: Kim in [1], proposed an iterative structure for deep neural networks to solve medical image segmentation task. In this work, an encoder-decoder architecture mixed with an iterative structure to improve the results of segmentation which consist of complex shapes in medical images. In addition, they utilized transfer-learning and data augmentation to prevent overfitting. Chang.Y in [2], proposed a method for cardiac medical image segmentation. Firstly, region of interest (ROI) detects by a YOLO structure then it feeds to a convolutional neural network for segmentation. Finally, the segmented images classify by a fully connected neural network and the physicians can detect the type of cardiac disease precisely. Isin in [3], discussed about different methods for brain tumor segmentation. These methods have been divided into three groups, including: non-automatic, semi-automatic and automatic methods. In automatic methods, there is no intervention from human and because of this, the probability of error decreases. Pim in [4], trained a CNN to segment six textures in brain MRI images, Chest muscle in muscle MRI images and vessels in cardiac MRI images. Therefore, it is a useful method for different organs segmentation and in different imaging conditions.

Dina Abdelhafiz in [5], proposed a U-net neural network to address the lesions in mammogram medical images precisely. In addition, they used data augmentation and batch normalization to improve the accuracy of their neural network. Hauzhu Fu in [6], segmented eye's vessels by a fully connected neural network. This network produces the feature maps of eye vessels, then they fed it to a CRF network to produce the binary images of eye's vessels. Zhe Gou in [7], analyzed medical images in different levels, including: feature learning, classifier level and decision making. Finally, they designed a neural network to detect malignant tumors in different imaging methods (CT, PET and MRI).

Celebi et al in [8], used a method for skin lesion detection by four threshold methods. Zhou et al in [9], segmented the skin lesion segmentation by mean shift estimation. This method needs many computational sequences. Xie et al in [10], combined a neural network method with genetic algorithm for skin lesion segmentation. Convolutional neural networks are one of the modern and practical methods in segmentation task, specifically for medical image segmentation [11]. Additionally, this method has been used for brain tumor segmentation in brain MRI images [12], cardiac image segmentation [13], skin lesion segmentation by non dermoscopic images [14] and many other medical images relating to the different parts of the body.

3. Methodology

Stochastic Gradient Descent (SGD) is the optimization algorithm that has been used in this method. It is an iterative method for optimizing an objective function with suitable smoothness properties. It can be regarded as a stochastic approximation of gradient descent optimization, since it replaces the actual

gradient by an estimate thereof. Specifically, in high dimensional optimization problems, it reduces the computational burden, achieving faster iterations in trade for a lower convergence rate. A good overview with convergence rates can be found in [15]. Nestrov momentum is a slight variation of normal gradient descent that can speed up training process and improve convergence significantly. Our goal is to learn a linear scoring function *f(x)*. In order to make predictions for binary classification, we simply look at the sign of *f(x)*. To find the model parameters, we minimize the regularized training error given by following formula:

$$\underline{\underline{E}}(w,b) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, f(x_i)) + \propto R(w) \tag{1}$$

In this formula, *(yi,f(xi))* is a sequence of training examples, L is a non-negative loss function and $\alpha > 0$ is regularization parameter. Two preprocess algorithms including image resizing and image augmentation have been done on the input images. Resizing decreases the dimensions of the input images that can speed up our implementation process. Also, image augmentation increases the number of input images that can increase the accuracy of our method and prevent overfitting. After these preprocessing algorithms, the images enter to the CNN for segmentation. The segmented output images are blurred, to solve this problem, we have proposed a postprocessing step. We used threshold technique as a postprocessing step to make images sharper. By this method, the physicians can detect the borders of lesions and tumors in medical images accurately.

3.1. Neural Network Architecture

The proposed neural network contains an encoder decoder structure that includes 64-layered network excluding the final activation layer. Every sequence of encoder has multiple convolutional layers, batch normalization with relu nonlinearity which is followed by non- overlapping max pooling and sub-sampling. At the center of the network, there are two dense layers that have been presented before the first up-sampling begins. The defining characteristic of proposed network is the use of max pooling indices in the decoders to perform up-sampling of low-resolution feature maps. This leads to keep of the important detailed features in the image and non-important features are erased. The architecture of proposed CNN has been shown in Figure1.

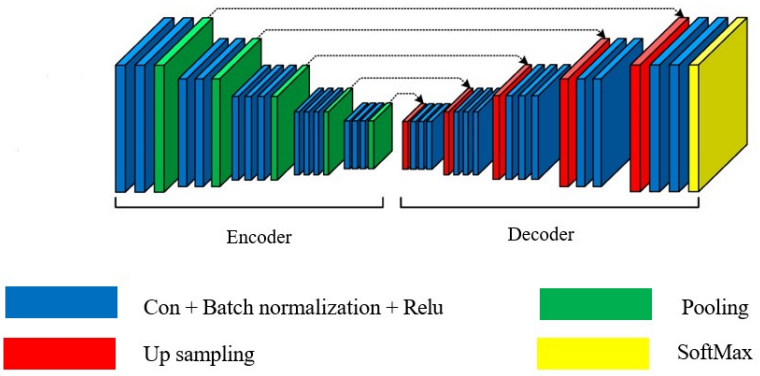

Figure 1. The Architecture of the Proposed CNN

The architecture consists of a sequence non-linear processing layers (encoders) and a corresponding set of decoders followed by a pixel-wise classifier. Typically, each encoder consists of one or more convolutional layers with batch normalization and a relu non-linearity, followed by non-overlapping max pooling and sub-sampling. The sparse encoding due to the pooling process is up-sampled in the decoder using the max-pooling indices in the encoding sequence. One key ingredient

of this method is the use of max-pooling indices in the decoders to perform up-sampling of low-resolution feature maps. This has the important advantages of retaining high frequency details in the segmented images and reducing the total number of trainable parameters in the decoders. The entire architecture can be trained end-to-end using SGD.

3.2. Loss Function

Crossentropy is function which measures how far away from the true value the prediction is for each of the classes and then averages the errors class wise to obtain the final lose. In this problem, there lies only two classes for each pixel, either black or white(0or1) as per the mask. So, here binary-crossentropy is used as the loss function rather than the categorical crossentropy originally proposed. The binary-crossentropy is in the below form:

$$H_p(q) = -\frac{1}{n}\sum_{i=1}^{N} y_i.\log\left(p(y_i)\right) + (1-y_i).\log\left(1-p(y_i)\right) \tag{2}$$

Where y is the label (1 for white points and 0 for black points) and *p(y)* is the predicted probability of the point being white for all N points. Reading this formula, tells us that, for each white point *(y=1)*, it adds *log(p(y))* to the loss, that is, the log probability of it being white. Conversely, it adds log *(1- p(y))*, that is, the log probability of it being black, for each black point *(y=0)*. Additionally, we have utilized Jaccard distance as loss function. This function has been defined as following formula [16]:

$$d_j(M,C) = 1 - J(M,C) = 1 - \frac{|M \cap C|}{|M| + |C| - |M \cap C|} \tag{3}$$

M represents the ground truth of segmentation, which is normally, a manually identified tumor region and *C* represents a computer-generated mask. *dj(M, C)* itself is not differentiable, which makes it difficult to be directly applied into backpropagation.

3.3. Network Training

For the training process, we utilized 75% of 200 images that are available in the PH2 dataset [17]. Because of image augmentation process as a preprocessing method, the number of input images increases to 450. 20% of these images associate to the validation dataset and the rest of them associate to the trainable dataset. As per the architecture of the network, the number of trainable parameters is 33,377,795 out of 33,393,669. Whereas, the non-trainable parameters are 15,874. The implementations are in the keras and the environment that has been used is google colab, the cloud space for python codes.

3.3. Image Augmentation

The image augmentation process that has been done in this method is for increasing the robustness of our method and reducing the chances of overfitting. Two techniques of image augmentation including rotation and horizontal flipping [18] have been done in this process. For rotation, the images are rotated around [-40,40] degrees and for horizontal flipping, the images are reversed around the horizontal axis. All the above transformations, are exactly performed on the corresponding masks of the images as well to maintain the correct orientation of feature images with their truth masks. As we have mentioned in the last part, after the augmentation, the transformed images are included in the training set which increases our training set from 150 to 450. Out of these 450 images, 90 images have been excluded to formation of a validation set and the rest of images associate to the training dataset. In the image augmentation technique, both original medical images and their corresponding masks have been affected by these techniques, resulted in, the number of images that are going to feed to the neural network, increases to 450. The results of our image augmentation technique have been displayed in Figure2.

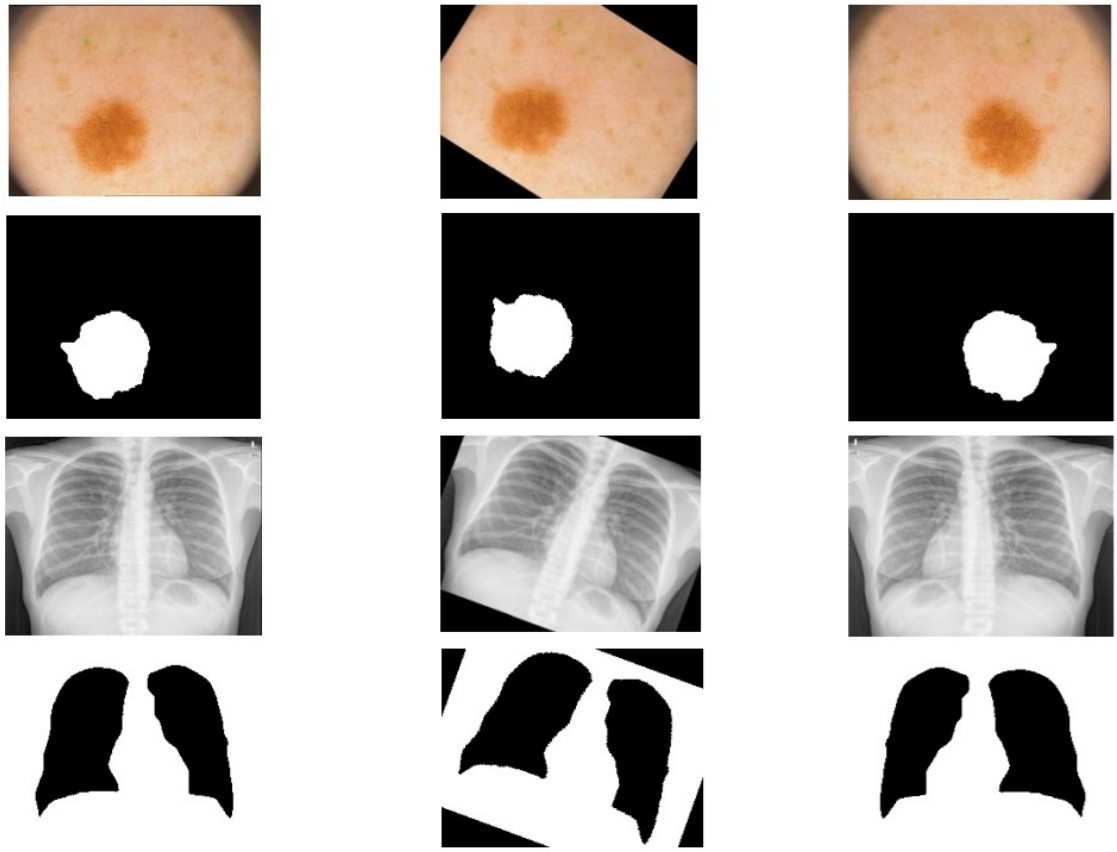

Figure2. Results of Image Augmentation:
(a) Original Medical Images with Their Corresponding Masks
(b) Medical Images After Rotation with Their Corresponding Masks
(c) Medical Images After Horizontal Flipping with Their Corresponding Masks

4. Finding and Discussion

4.1. Datasets

The used datasets in this paper, are medical images dataset which includes dermoscopic and Chest-Xray images. The number of images in each is 200 for original medical images and 200 for their corresponding masks. The dimensions of dermoscopic images are 572*765 and the dimensions of Chest-Xray images is 512*512 before resizing. Both datasets have been publicly available for programmers. The PH2 is acquired at the dermatology service of hospital Pedro Hispano, Matosinhos, Portugal [19] and the chest-Xray medical images have been provided in Kaggle website [20].

4.2. Performance Evaluation

The outputs of proposed CNN are binarized into lesion masks. The performance of the CNN has been assessed by comparison between the neural network generated lesion masks and the masks of medical images which are provided in medical image dataset. We have assessed the performance of our method by the following metrics:

- **Jaccard Index**

Jaccard Index (JA) also known as Intersection Over Union (IOU). The Jaccard similarity coefficient is a statistical similarity measure to check the diversity among the sample sets. The IOU gives the similarity among sets and the formula is the size of the intersection over the size of the union of the sets. We can define evaluation metrics based on confusion matrix components. Let TP, TN, FP, FN

refer to the number of true positives, true negatives, false positives, and false negatives. We can show Jaccard distance by JA and define it as following formula:

$$JA = \frac{TP}{TP + FN + FP} \tag{4}$$

- **Dice Coefficient**

The Dice score (F1 score) is like precision. It measures the positives as well as it applies penalty to the false positives given by the model. It is more similar to precision than accuracy. We have shown dice coefficient by DI and have calculated it by the following formula:

$$DI = \frac{2 * TP}{(TP + FP) + (TP + FN)} \tag{5}$$

- **Recall**

Recall is a measure which is targeted towards the actual or the true positives yielded by the model output. In the scenarios where the cost of the False Negatives is greater than recall is the better metric to choose the best model among the possible ones. Recall also known as sensitivity (SE). we can calculate Recall as following formula:

$$SE = \frac{TP}{TP + FN} \tag{6}$$

- **Accuracy**

Accuracy approximates that how close our predicted output to the real ground truth. We have shown it by AC and have calculated it by the following formula:

$$AC = \frac{TP + TN}{TP + FN + TN + FP} \tag{7}$$

4.3. Comparison with Two Famous Architectures

We compared our results with two famous architectures, including Unet and FCN neural networks. We show that our method could outperform these two architectures in medical image segmentation. FCN, despite up-convolutional layers and a few shortcut connections produces coarse segmentation maps. Therefore, more shortcut connections are introduced. However, instead of copying the encoder features as in FCN, indices from max-pooling are copied. This proposed method more memory-efficient than FCN. Additionally, in proposed method only the pooling indices are transferred to the expansion path from the compression path, using less memory. Whereas in the Unet, entire feature maps are transferred from compression path to expansion path making, using a lot of memory. We have shown our results in Table 1.

Table 1. The Comparison between the Results of Our Method and the Results of Other Methods for PH2 Dataset

Method	Optimization Algorithm	AC	DI	JA	SE
Yuan [16]	SGD+Nestrov	0.954	0.907	0.838	0.910
	Adam	0.963	0.922	0.861	0.926
FCN	SGD+Nestrov	0.917	0.648	0.878	0.822
	Adam	0.870	0.694	0.876	0.878
Unet	SGD+Nestrov	0.842	0.644	0.866	0.561
	Adam	0.909	0.853	0.921	0.930
Proposed Method	SGD+Nestrov	0.972	0.765	0.945	0.963
	Adam	0.884	0.762	0.878	0.988

Table 2. The Comparison between the Results of Our Method with the Results of Another Methods, based on Different Loss Functions for PH2 Dataset

Method	Optimization Algorithm	AC	DI	JA	SE
Yuan [16]	SGD+Nestrov	0.963	0.922	0.861	0.926
	Adam	0.961	0.916	0.854	0.918
FCN	SGD+Nestrov	0.917	0.648	0.878	0.822
	Adam	0.934	0.692	0.902	0.867
Unet	SGD+Nestrov	0.909	0.853	0.921	0.930
	Adam	0.939	0.840	0.945	0.858
Proposed Method	SGD+Nestrov	0.972	0.765	0.945	0.963
	Adam	0.975	0.839	0.965	0.970

As shown in the Table 1, our method has better results than other methods with SGD+Nestrov optimization algorithm, except dice coefficient (DI). We can compare our results with other methods, based on different loss functions. We have shown our results in the Table 2.

As shown in the Table 2, except dice coefficient, our method has better results than the other algorithms based on two jaccard distance and crossentropy loss functions. We have compared our method with two Unet and FCN methods, because they have robust results in medical image segmentation, specifically in the mentioned datasets, and many works have been done by these two methods. Additionally, we have compared our method with other architectures and the method that has implemented in [21] with SGD optimization for Chest-Xray dataset. The results of our comparison have been shown in Table 3.

Table 3. The Comparison between the Results of Our Method with the Results of Another Methods, based on SGD Optimization Algorithm for Chest-Xray Images

Method	Optimization Algorithm	AC	DI	JA	SE
Jonathan [21]	SGD+Nestrov	0.969	0.935	0.880	0.975
FCN	SGD+Nestrov	0.890	0.894	0.909	0.868
Unet	SGD+Nestrov	0.911	0.870	0.929	0.983
Proposed Method	SGD+Nestrov	0.989	0.933	0.981	0.984

As shown in the Table 3, our method has a better performance in comparison to other methods based on SGD optimization algorithm for Chest-Xray images. The evaluation metrics report that our architecture is better than the others for medical image segmentation.

4.4. Simulation Results

We have displayed the results of FCN and Unet neural networks in Figure3. The borders between Region of Interests (ROI) and other regions in final segmented medical images, are not very clear.

Additionally, we have displayed the results of medical image segmentation that has implemented by our method in Figure4. As it shows, the borders between ROI and other regions are very clear that physicians can detect the lesions or Chest problems precisely.

Therefore, our method is a robust and precise method toward medical image segmentation task. The Figure4 shows that the borders between ROI and other regions are more clear and more precise than the borders in Figure3. All of these medical images have the mentioned problems like irregular and fuzzy borders, low contrast between the lesions and other surrounding textures, hair artifacts and etc. Our method could tackle with all of these challenging features.

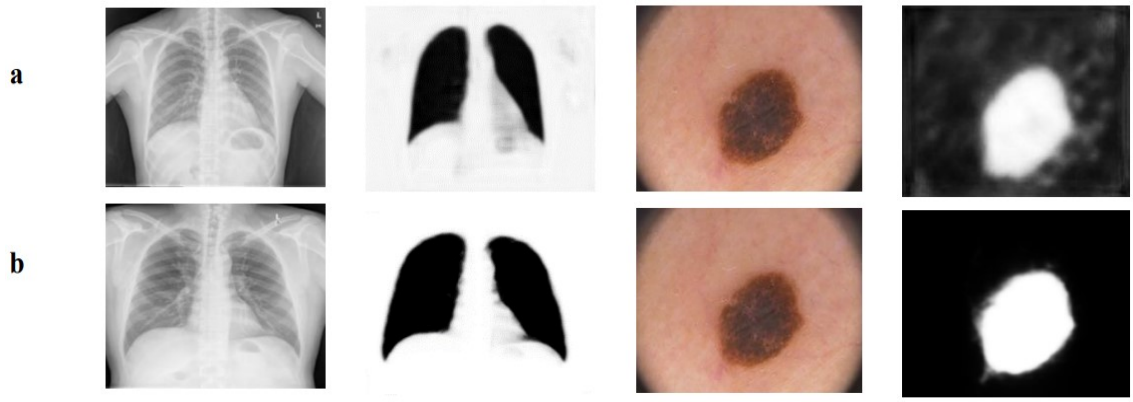

Figure 3. The Results of FCN and Unet Neural Networks:
(a) Medical Images with Their Corresponding FCN Segmentation
(b) Medical Images with Their Corresponding Unet Segmentation

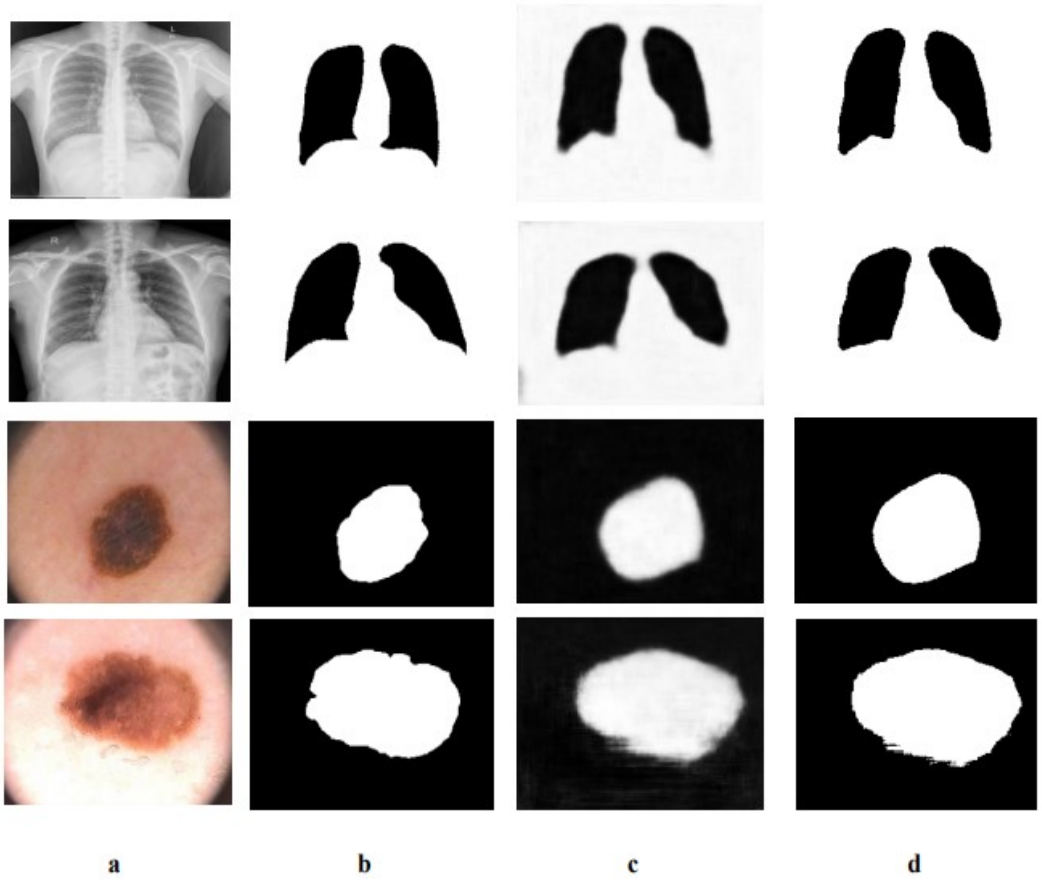

Figure 4. The Results of The Proposed Method:
(a) Original Medical Images
(b) Corresponding Masks of Medical Images
(c) Results Of Segmentation by The Proposed Method
(d) Final Results After Postprocessing

5. Conclusion

In this method, we rendered a novel architecture toward medical image segmentation task. Several effective training strategies were implemented to tackle the challenges that training a CNN may face when only limited training data is available. We designed our loss function based on both Jaccard distance and binary-crossentropy. Compared to the conventionally used cross entropy, the Jaccard distance-based loss function directly maximizes the overlap between the foreground of the ground truth and that of the predicted segmentation mask, and thus eliminates the needs of data re balancing when the numbers of foreground and background pixels are highly unbalanced, such as binary medical image segmentation. Two techniques of image augmentation, image rotation and horizontal flipping on the training dataset are performed before feeding it to the network for training. After the training process the model was evaluated on several measures for statistical values. The predictions produced from the model on test images were post-processed using the thresholding technique to remove the blurry boundaries around the predicted lesions. We compared the results of our method with the results of two other state of the art architectures including Unet and FCN. The results show that our method could outperform these two famous and practical architectures in medical image segmentation task. We hope in future works, will develop our method for increasing the dice coefficient and implement it on other famous datasets, like: ISIC2017[22] and ISBI2016[23].

Acknowledgment

The authors are thankful to the dermatology service of hospital Pedro Hispano in Portugal, where makes the 11 dermoscopic image data available. They also thank the anonymous reviewers whose comments and suggestions helped improved this manuscript. Additionally, we are thankful to the Kaggle website, where has provided so many different datasets for researches and programmers.

References

[1] Kim, J.U., H.G. Kim, and Y.M. Ro. Iterative deep convolutional encoder-decoder network for medical image segmentation. in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017. IEEE.

[2] Chang, Y., et al. Automatic Segmentation and Cardiopathy Classification in Cardiac Mri Images Based on Deep Neural Networks. in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018. IEEE.

[3] Işın, A., C. Direkoğlu, and M. Şah, Review of MRI-based brain tumor image segmentation using deep learning methods. Procedia Computer Science, 2016. 102: p. 317- 324.

[4] Moeskops, P., et al. Deep learning for multi-task medical image segmentation in multiple modalities. in International Conference on Medical Image Computing and Computer- Assisted Intervention. 2016. Springer.

[5] Abdelhafiz, D., et al. Convolutional Neural Network for Automated Mass Segmentation in Mammography. in 2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS). 2018. IEEE.

[6] Fu, H., et al. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. in 2016 IEEE 13th international symposium on biomedical imaging (ISBI). 2016. IEEE.

[7] Guo, Z., et al., Deep Learning-Based Image Segmentation on Multimodal Medical Imaging. IEEE Transactions on Radiation and Plasma Medical Sciences, 2019. 3(2): p. 162- 169.

[8] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, “Lesion border detection in dermoscopy images usingensembles of thresholding methods,” Skin Res. Technol., vol. 19, no. 1, pp. e252– e258, 2013.

[9] H. Zhou, X. Li, G. Schaefer, M. E. Celebi, and P. Miller, “Meanshift based gradient vector flow for image segmentation,” Comput.Vis. Image Understand., vol. 117, no. 9, pp. 1004–1016, 2013.

[10] F. Xie and A. C. Bovik, “Automatic segmentation of dermoscopy images using self- generating neural networks seeded by genetic algorithm,” Pattern Recognition, vol. 46, no. 3, pp. 1012–1019, 2013.

[11] Mendonça, T., et al. PH 2-A dermoscopic image database for research and benchmarking. in 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC). 2013. IEEE.
[12] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," IEEE Trans.Med. Imag., vol. 35, no. 5, pp. 1240–1251, 2016.
[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradientbased learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
[14] S. Ioffe, C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,11February 2015. [Online].Available:https://arxiv.org/ abs/1502.0316 7. [Accessed July 2019].
[15] T. Zhang "Solving large scale linear prediction problems using stochastic gradient descent algorithms" - In Proceedings of ICML '04.
[16] Y. Yuan, M. Chao, Y.C. Lo, "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance,"IEEE Transactions on Medical Imaging, 2017.
[17] M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. Soroushmehr, S. Samavi, and K. Najarian, "Extraction of skin lesions from nondermoscopic images using deep learning," arXiv preprint arXiv:1609.02374, 2016.
[18] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deeplearning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," Med. Image Anal., vol. 30, pp. 108–119, 2016.
[19] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, J. Rozeira, "PH2 - A dermoscopic image database for research and benchmarking," 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5437 - 5440, 2013.
[20] https://www.kaggle.com/search?q=Chest+xray+segmentation.
[21] Johnatan Carvalho Souza, Joao Ot ˜ avio Bandeira Diniz, ´ Jonnison Lima Ferreira, Giovanni Lucca Franc¸a da Silva, Aristofanes Corr ´ ea Silva, ˆ Anselmo Cardoso de Paiva, An automatic method for lung segmentation and reconstruction in chest X-Ray using deep neural networks, Computer Methods and Programs in Biomedicine (2019).
[22] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A. Halpern, "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)," 2017. [Online]. Available: https://challenge.kitware.com/#challenge/583f126bcad3a51cc66c8d9a.
[23] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A. Halpern, "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)," May 2016. [Online]. Available: https://arxiv.org/abs/1605.01397.