Melanoma Segmentation and Classification using Mask R-CNN and AlexNet

1st Darren Shih
Morrison Academy Taichung
Taichung, Taiwan
shihd@mca.org.tw

2nd Chia-Shin Wei
Medical Informatics
Chung-Shan Medical University
Taichung, Taiwan
sabrina0226wei@gmail.com

3rd Ming-Chen Hsu
Medical Informatics
Chung-Shan Medical University
Taichung, Taiwan
exo050282@gmail.com

4th Chiun-Li Chin
Medical Informatics
Chung-Shan Medical University
Taichung, Taiwan
ernestli@csmu.edu.tw

***Abstract*—It is crucial for Melanoma, a type of skin cancer, to be diagnosed early. Many have used the The International Skin Imaging Collaboration(ISIC) archive to train their programs to better recognize it. However, studies have shown that the surrounding skin may affect the classifying performance, leading to biased results. This might imply lower classification rates in the real world. This paper combines both segmentation and classification. We apply Mask R-CNN for firstly segmenting out the skin lesion from original images. The segmented images are then passed into an AlexNet architecture which then classifies the lesion. From the results of this research, the precision of segmentation and classification can reach a high rate of 92.6% and 87.79% on average respectively. It shows that the model could achieve great performance in segmenting the regions of the melanoma, and further proves that the following classifying model, AlexNet, could have positively great performance in classification without any other negative influence caused by the surrounding skin part. Then the result is potentially accurate enough to be considered as great assistance in real world clinical practice. In the future, we hope that there are more types of skin lesions that could be identified by our model. By the means of more widely recognized skin lesions, our research could assist doctors with comprehensive diagnosis, which would have optimistic improvement towards clinical application.**

Keywords—melanoma, segmentation, classification, Mask R-CNN, AlexNet

I. INTRODUCTION

Melanoma is the most deadly form of skin cancer. Despite it consisting of only 1% of diagnosed skin cancer in the US, melanoma causes the most deaths compared to other types. If melanoma was diagnosed within five years, survival rates are high, some 92%. After five years, if melanoma has spread to other parts of the body, the survival rate drops to 25% [1]. Therefore, it is crucial that melanoma is diagnosed early. The most direct way to detect melanoma is to observe skin lesions from the images. Normally, the skin lesions are present in two types, benign and malignant, which are shown as Fig.1. Nowadays, while doctors judge the melanoma on the images whether belonging to benign or malignant, sometimes they could not distinguish because the appearance or color is too close, leading to a lot of time spent in diagnosis and treatment, and that the patients also have to wait a long time. To address this issue, the use of deep learning methods was employed including architectures of deep neural networks.

(a) (b)

Fig. 1. (a) Benign image, (b) melanoma image.

Bissoto, A., Fornaciali, M., Valle, E., and Avila, S. [2] explored the bias within the commonly used skin lesion datasets through exploring deep learning algorithms' ability to learn even when the lesion is blocked. They tested the algorithm through a variety of different levels of blocking from blocking the lesion around the border, blocking the lesion in a box, and blocking 70% of the image that contained the lesion. Through the use of 10 splits and cross-dataset, they sought to eliminate bias within the databases. However, they discovered that the algorithms were still able to recognize between malignant and benign lesions, a solid 70%, even when most of the lesion was blocked. Then, they attempted to guide the network's learning process by blocking out the surrounding area and giving it grayscale image, RGB attributes image, and the combination of the two. However, the images were not cut out cleanly. The result did not improve the model and resulted in a lower accuracy than the training of the original images.

Yu, L., Chen, H., Dou, Q., Qin, J., and Heng, P.-A. [3] used a very deep neural network of more than 50 layers to acquire more features, while using residual learning to prevent degradation and overfitting. Then they segment the skin lesion through the use of fully convolutional residual network (FCRN). Passing the cropped image (length and width) into the second stage of the framework, the use of another deep residual network was utilized to classify the image. They got second out of the 28 teams that participated in the segmentation challenge on ISBI 2016 and got first in the skin lesion classification challenge on ISBI 2016. Li, Y. and Shen, L. [4] also used two deep learning methods to do segmentation and classification through FCRN. This algorithm was validated on the ISIC 2017 dataset and achieved a high accuracy. However, in both papers, the rectangular-cropped image passed also consisted of the surrounding skin, which may have been redundant or biased.

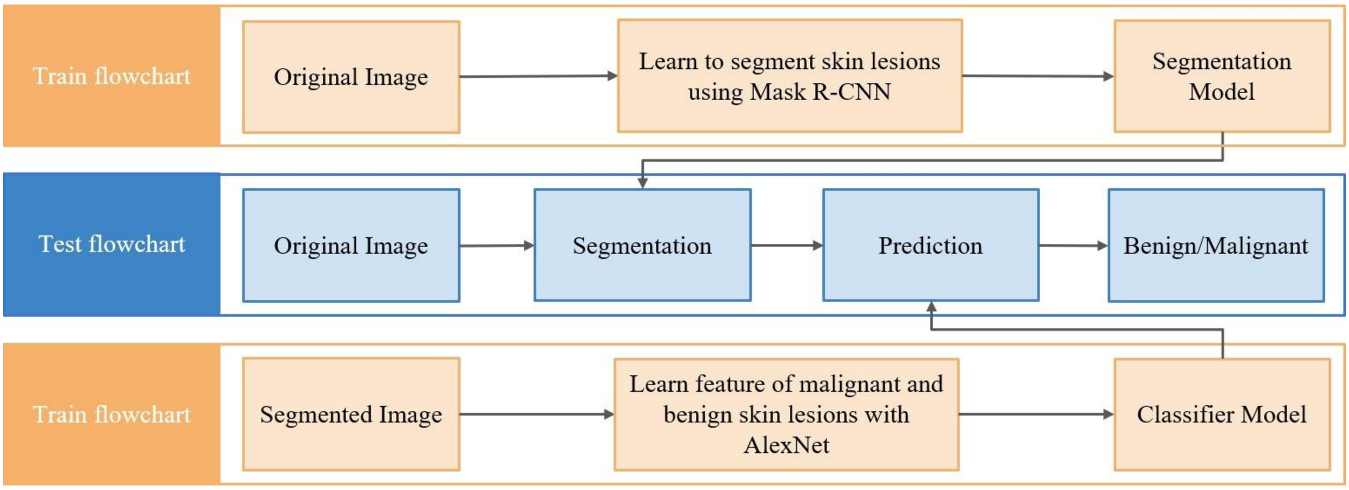


Fig. 2. Flowchart of the system

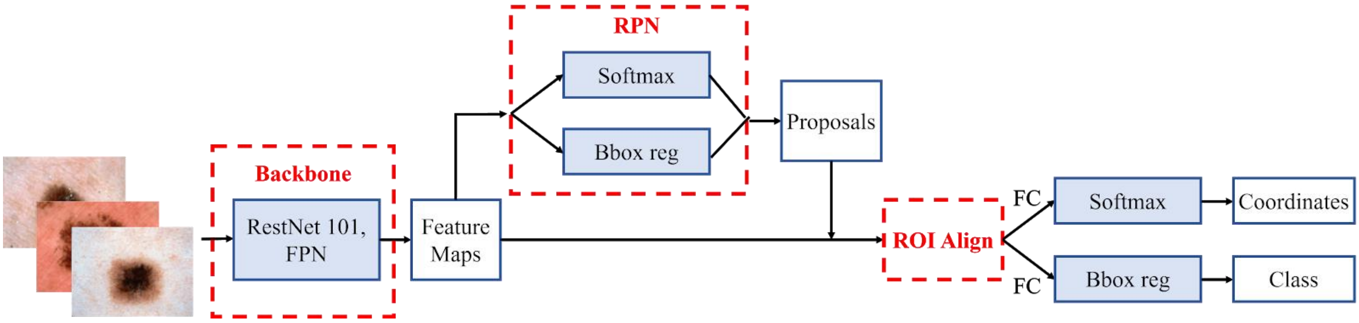


Fig. 3. The architecture of the Mask R-CNN network

Many scholars have devoted themselves to the study of lesions segmentation. In 2017 Xiaoqing Z. [5] and in 2019 Eric Z. Chen et al. [6] utilized the U-Net deep learning architecture to accomplish the task of the melanoma segmentation. Through the architecture, the feature map will be combined with the features extracted again in the next layer to generate the segmenting area of the melanoma . Although this method can achieve good performance in segmenting, due to the method of U-Net architecture in segmenting the melanoma area is to combine the feature map, thus to use bilinear interpolation for segmenting instead of the original method.

Training primarily on public databases of skin lesions, many neural networks have achieved a high accuracy rate. However, the fact that they may have learned features from the surrounding skin poses an issue of bias. The algorithm may obtain good results from a biased source and does not perform well in the real world application of classifying the skin lesion. The papers above are taken as reference and lead into the custom method used in this paper.

II. METHOD AND RESULTS

In this paper, we use a custom method that combines both segmentation and classification, the flowchart of the system is shown as Fig.2.

A. *Dataset*

The training and testing dataset that we utilized in this research is from the International Skin Imaging Collaboration (ISIC), which contains benign and malignant. We use 1504 images for training, and 376 images as the testing dataset.

B. *Mask R-CNN*

A Mask R-CNN model is trained to recognize the regions of skin lesions and applies masks on the input images to segment out the lesion area. The architecture of the Mask R-CNN network we utilize is shown in Fig. 3. Mask R-CNN, which is the state-of-the-art instance segmentation deep learning model, is mainly composed of two parts of architecture, including before obtaining the proposal and after obtaining the proposal. The part before obtaining proposals is assembled by a backbone network and the Region Proposal Network (RPN). The backbone network consists of the ResNet-101 and Feature Pyramid Network (FPN), which is utilized for extracting subtle features from images and produces the feature maps. The output feature maps would then be input into the RPN to detect possible objects of melanoma as well as to obtain respective coordinates of each object, which is inferred as generating proposals. After the procedure of capturing proposals, the proposals and its corresponding feature maps will then be fine-tuned through RoI Align to produce the final proposals. In the next step, the proposals will go through a fully connected (FC) layer to perform pixel-wise classification and coordinate correction. Eventually, the Mask R-CNN model will output the objects with class and its coordinates.

Above is the overall progress of what the Mask R-CNN model does, in the following, we utilize a self-built algorithm to segment out the identified objects, which keeps original.

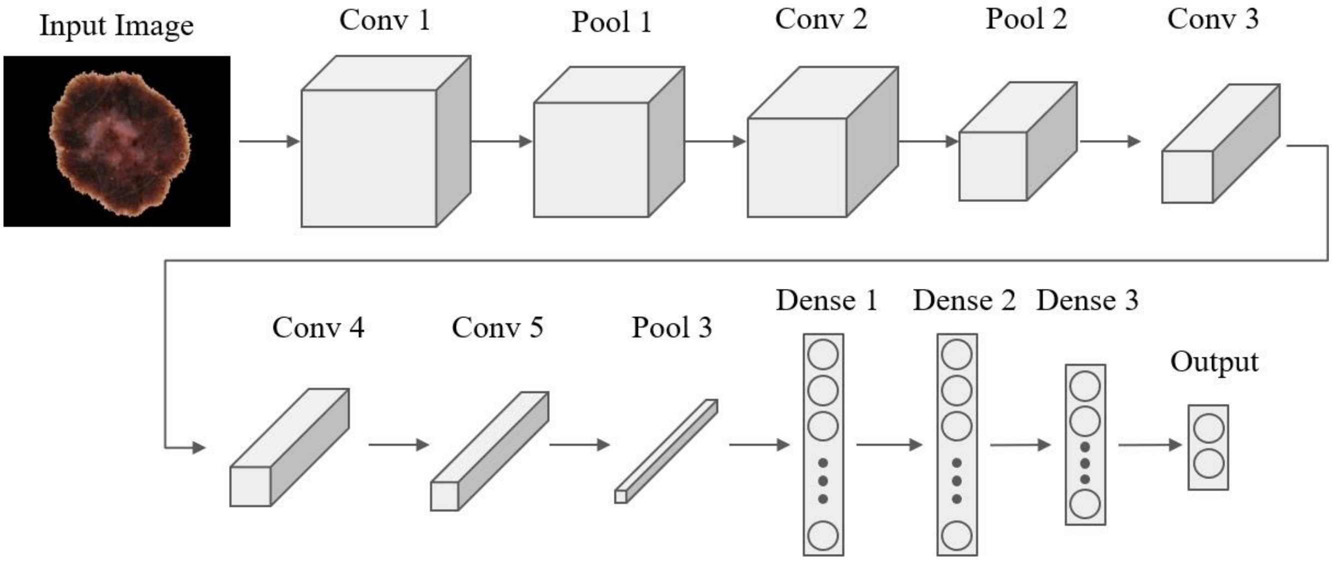
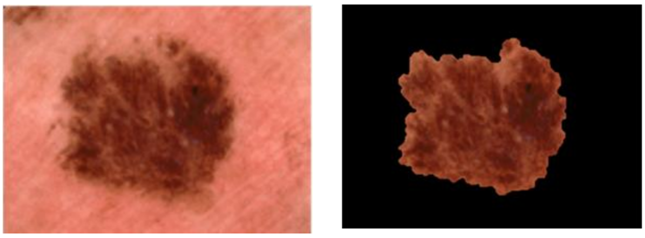


Fig. 4. Classifier Model: AlexNet architecture



(a) (b)

Fig. 5. (a) is the skin lesion image, (b) is the segmentation of skin lesion image

colors of skin lesion while replacing the surrounding parts into a black color. The skin lesion image is shown as Fig.5 (a), the segmentation of skin lesion image is shown as Fig.5. (b). Blocking out the skin surrounding the lesion allows the next classification algorithm to learn the correct features of a skin lesion and also eliminates redundancy of the repeating skin.

C. Modified AlexNet

Each segmented image is passed into a modified AlexNet network, shown as Fig. 4, which then classifies the lesion. The AlexNet architecture was modified to have the number of neurons in each dense layer to equal 1000, 1000, 500, and 2. A dropout of 0.6 was added to prevent overfitting. The loss function we adopted is binary cross-entropy due to the classification of two different types: melanoma and nevus. The optimizer utilized is Stochastic Gradient Descent (SGD) which finds the appropriate value of weights of the network. To further prevent the bias potentially obtained in the database, we apply 5-fold cross-validation which split the dataset into 5 different sections, and 4 were randomly selected to train while the last one was used for validating the testing result.

III. Experimental Results

In the segmentation of melanoma with Mask R-CNN, Mask R-CNN model could achieve excellent performance in testing. Table I. shows the testing confusion matrix of Mask R-CNN. Additionally, due to the high similarity within the lesion parts against the surrounding skin, despite the class of nevus, we take background as one of the classes into the confusion matrix for evaluating the error rate of segmentation failure. Among a total of 376 test images, the segmentation accuracy of the Mask R-CNN model can reach 98.67%, and the probability that the background is misjudged as nevus is 0.808%.

Even though Mask R-CNN has a great segmenting ability, there are still defects that we found among the segmentation. The mistake of segmentation by Mask R-CNN is shown in Fig. 6. We determined this defect of segmentation is due to the ambiguous contour of nevus against the surrounding skin tissue, which leads to the misjudgment of segmenting.

(a) (b)

Fig. 6. (a) Original image of nevus, (b) The nevus segmentation mistaken

After performing the segmentation by Mask R-CNN, the segmented objects are then taken into our modified AlexNet to identify which class the object belongs to. To evaluate the classification ability of the modified AlexNet model, four different aspects are measured to determine the proficiency of the model: Accuracy, Sensitivity, Specificity, and Precision which are 80.16%, 70.21%, 90.11%, and 87.79% respectively.

TABLE I. Mask R-CNN Testing Confusion Matrix

Confusion Matrix		Predicted Class	
		Nevus	Background
Actual Class	**Nevus**	177	11
	Background	13	-

According to the confusion matrices in Table I and Table II, we thus can refer that the model could significantly achieve great performance in both segmentation and classification tasks.

TABLE II. AlexNet Testing Confusion Matrix

Confusion Matrix		Predicted Class	
		Nevus	Background
Actual Class	**Nevus**	138	50
	Background	16	172

IV. Conclusion

In this paper, we propose a method that combines both segmentation with Mask R-CNN and classification with a modified AlexNet that is used to identify malignant melanoma from benign nevus. Despite their similarities, the proposed model is effective in recognizing the differences, achieving accurate results. For application purposes, this model can be used to assist a doctor in clinical practice examinations of a skin lesion, or possibly this model can be used as a check for those who are suspicious of their own skin condition. Further investigations include using another model to either segment skin the skin lesion or classify the cropped skin lesion.

V. Acknowledgments

In this research, we especially appreciate the work of the International Skin Imaging Collaboration (ISIC) for providing the organized public database of skin lesion images.

VI. References

[1] American Cancer Society's (ACS) publication, Cancer Facts & Figures 2020, and the ACS website (January 2020)

[2] A. Bissoto, M. Fornaciali, E. Valle, and S. Avila. "(De) Constructing bias on skin lesion datasets," IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.

[3] Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A., 2017a. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging 36 (4), 994–1004. doi:10.1109/TMI.2016.2642839.

[4] Y. Li, L. Shen, "Skin Lesion Analysis Towards Melanoma Detection Using Deep Learning Network," International Skin Imaging Collaboration (ISIC) 2017 Challenge at the International Symposium on Biomedical Imaging (ISBI), Available: https://arxiv.org/abs/1703.00577

[5]]X. Zhang, "Melanoma segmentation based on deep learning," Computer Assisted Surgery, vol. 22, pp. 267-277, 2017.

[6] E. Z. Chen, X. Dong, X. Li, H. Jiang, R. Rong and J. Wu, "Lesion Attributes Segmentation for Melanoma Detection with Multi-Task U-Net," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 485-488, 2019.