Skin Lesion Segmentation via Deep Learning

Kritika Pasricha
Center for Biomedical Engineering
Indian Institute of Technology Ropar
Rupnagar, India
2019bmm1007@iitrpr.ac.in

Roshni Jhariya
Center for Biomedical Engineering
Indian Institute of Technology Ropar
Rupnagar, India
2019bmm1011@iitrpr.ac.in

***Abstract*—Background and objective: Skin Lesion Segmentation can be challenging as size and shape of lesions vary largely. Digital analysis of skin lesion images had been commonly used by dermatologist for detection of melanoma. Computer assisted Segmentation of skin lesion help dermatologist for early detection and diagnostic of cancer.**
Methods: In this study we implemented deep learning architecture SegNet, UNet, and FCN (Fully Convolutional Neural Network) for skin lesion segmentation. These Deep Learning architectures have included some additional features for improvisations of results. Models were trained and tested on ISIC 2016 challenge dataset and evaluated with Intersection over Union (Jaccard Index), Dice coefficient, Precision, Recall and Accuracy.
Results: UNet segmentation network achieved highest Dice coefficient value of 89.66 and considered best among 3 proposed models. It achieved accuracy of 94.81%.

Keywords—Skin Lesion, Segmentation, Deep Learning, SegNet, UNet, Fully Convolutional neural network, Jaccard Index, Dice coefficient, precision, sensitivity.

I. INTRODUCTION

A. Identified Problem and Data Source

Melanoma skin cancer is the most frequently reported skin cancer type. Early symptoms of melanoma are different type of skin lesions like irregular spots, moles, sore, pigmentation, red patch or nodules over skin, varying size of globules and dots, streaking projection. These lesions have poorly defined border, asymmetric pigmentation, varying shades from black, brown to red and pink [1]. Skin lesion examination using Dermatoscope is a common practice [2]. Dermatoscopy allows deeper layer analysis with 75-84% accuracy. Altered pigmented surface are diagnostic feature of melanoma skin cancer which are identified through dermatoscopy, a noninvasive imaging technique. Segmentation of image captured from dermatoscope is vital for characterization of skin lesion for diagnostic purpose. It differentiates the abnormal skin texture from normal skin texture [3]. Manual segmentation requires expertise and pretraining due to large variation in skin lesions. It is also time consuming process. Early diagnostic is important to prevent spread of disease and timely treatment that increases the recovery rate. Various computer aided automatic image analysis method are mentioned in literature for evaluation of lesions [4]. Efficient segmentation of captured image is vital for any automatic computer aided analysis. Classical method for segmentation are nott efficient due to ambiguity in morphology of lesion. Various deep learning methods for skin lesion segmentation are discussed in literature. Fully automated skin lesion segmentation is challenging due to large variation in color, texture and undefined boundaries [3]. This work aimed to develop deep learning models to improve skin lesion segmentation using publicly available database, the IEEE International Symposium on Biomedical Imaging (ISBI) 2016 Challenge. The International Skin Imaging Collaboration (ISIC) provides a data of clinical and dermatoscopic images of skin lesion publicly to support their aim to reduce mortality related to skin cancer [5]. The challenge hosted by ISIC in the early 2016.

B. Existing solutions

Segmentation of skin lesion is prerequisite for any automated skin lesion detection algorithm. Computer aided digital analysis method had been long proposed. Computer aided digital dermoscopic image analyzer is popularly used to aid screening of skin lesion which provide sensitivity upto 56% [4], which makes it unreliable diagnostic tool. Computer aided Multispectral digital skin lesion analysis (MSDSLA) technology reported by Winkelmann RR et al. is also available, which uses 10 bands visible and near infrared light to evaluate skin lesion [6]. This technology help in reducing the unnecessary biopsies and provided accuracy upto 83% [6]. Several work has also been reported in literature to differentiate skin lesion region in input images. Garnavi et al. study on lesion border detection using optimal color space channels and clustering based thresholding technique for skin lesion analysis [7]. Ionut et al. developed active contour based segmentation method for macroscopic pigmented skin lesion images captured using smart phones in uncontrolled environment [8]. Celebi et al. worked on unsupervised approach to border detection method using statistical region merging algorithm [9]. Deep learning methods like convolutional neural network (CNN) are recently introduced for segmentation to outperform the performance of traditional image processing methods. Jane et al. investigated three different models for skin lesion segmentation, model 1 combined pre-trained VGG16 with U-net, model 2 used SegNet instead of U-net, and model 3 used DeeplabV3 [10]. Semantic segmentation network consisting of encoder and decoder block is also reported in literature for pixel wise segmentation with 92% accuracy [11]. A. Adegun et al. performed comparative analysis of two Fully Convolutional Encoder-Decoder Architecture, viz. U-net and SegNet for skin lesion segmentation [12]. Plenty of work on Deep leaning algorithm had been proposed and evaluated in literature for their applicability in detection of Skin lesion boundary. But due to large ambiguity in the morphology of pigmented lesions, it is still challenging task.

II. MATERIAL AND METHODS

A. *Exploratory Data Analysis*

The ISBI 2016 challenge dataset called “Skin Lesion Analysis Towards Melanoma Detection” contains original image, paired with the expert manual tracing of the lesion boundaries in the form of a binary mask. The training dataset provides 900 dermoscopic lesion images and test dataset provides 379 images available in JPEG format. All images are named using the scheme “ISIC image id.jpg”, where “image id” is a 7- digit unique identifier. The training ground truth file contains 900 binary mask images in PNG format. All masks are named using the scheme “ISIC image id Segmentation.png”. Mask images are encoded as single-channel (grayscale) 8-bit PNGs, where each pixel is either 0 (area outside lesion) or 255 (area inside lesion). Binary mask were created using either semi-automated method or manual process by opinions of expert dermatologist. The dataset contains mix images of both malignant and benign skin lesions. Also the training data will be divided into 80:20 as training data and validation data. Image augmentation was done on training dataset to boost the performance of model and reduce the chance of over fitting [13].

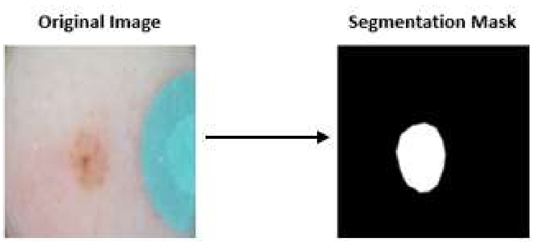

Fig. 1: Skin Lesion Segmentation [5]

B. *Data pre-processing and Augmentation*

The data images pixel values were from 0-255, they were normalized to 0-1, as pixels must be labelled as 0 for background (black) or 1 for lesion area (white). A function named ’binary’ was used to binarize the image dataset and the images were resized to 320x320 pixels For Segnet Model and FCN model, 224x224 pixel for UNet model. Data is numerically sorted, so that original image and its corresponding masks can be synchronized. After this data is appended in a list and then the list is converted into a NUMPY array, as models accepts data in array form. Increasing data size increases the performance of the model. So we increased the data size to 2160 images for training dataset and 379 images for test dataset and 180 images for validation dataset. For augmentation purposes the data images are horizontally flipped and randomly rotated in between -40 to 40 degree angles.

Table 1: ISIC 2016 Dataset

Number of images	Before Augmentation	After Augmentation
Training	720	2160
Test	379	379
Validation	180	180

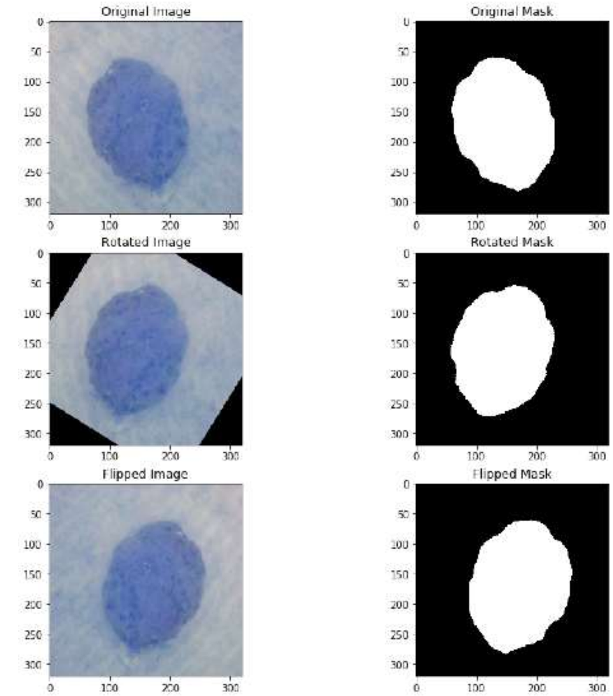

Fig. 2: Images after Augmentation

C. *Proposed Machine learning models*

We investigate the applicability of deep learning methods for skin lesion segmentation. We evaluated performance of three architectures: SegNet, U-Net and Fully Convolutional neural network (FCN).

SegNet: This architecture is popular for its efficient pixel wise semantic segmentation [14]. It is based on convolutional encoder-decoders layers for semantic pixel wise labelling [13]. Network Architecture built is 90-layered network including the final activation layer. Every sequence of encoder has multiple convolutional layers, batch normalized with ReLU nonlinearity which is followed by non-overlapping maxpooling and sub-sampling. At the center of the network there are two dense layers present before the first up-sampling begins. Optimizer and Learning Rate- We adopted SGD optimization algorithm, to adjust the learning rate. It is well known that learning rate is one of the critical hyper-parameters that have a significant impact on classification performance. On large dataset, SGD can converge faster than batch training because it performs updates more frequently. We can get away with this because the data often contains redundant information, so the gradient can be reasonably approximated without using the full dataset. Minibatch training can be faster than training on single data points because it can take advantage of vectored operations to process the entire minibatch at once. The stochastic nature of online/minibatch training can also make it possible to hop out of local minima that might otherwise trap batch training. The learning rate is the conventional 0.001. Images were resized to 320 x 320 px. We used sigmoid as final activation function as using softmax was producing blank dark images. The loss function in SegNet is suggested categorical crossentropy due to its general use of multiclass segmentation problem but in our case we only have 1 class i.e. the melanoma. So, we used binary crossentropy.

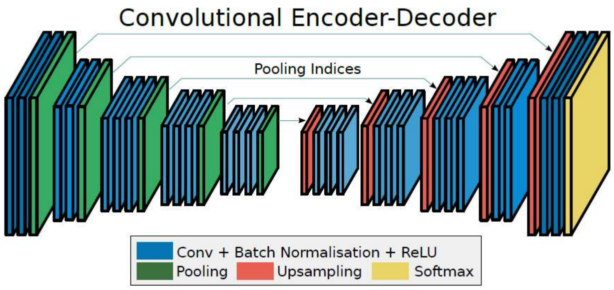

Fig. 3: SegNet Architecture [14]

UNet: UNet is end to end efficient fully convolutional network which relies on strong use of data augmentation (15). Images were resized to 224 x 224 px. The network consists of the repeated application of two 2x2 convolutions with the same number of feature maps, we have used batch normalization and spatial dropout in our network which showed improved performance. Rectifier Linear Unit (ReLU) was used as activation function and 2×2 max pooling was applied for downsampling. In the latter half of the network, for upsampling of the feature maps, a 2×2 up-convolution is applied followed by feature concatenation and two 2x2 convolutions. At the final layer, 1×1 convolution is applied and sigmoid activation is used [15].

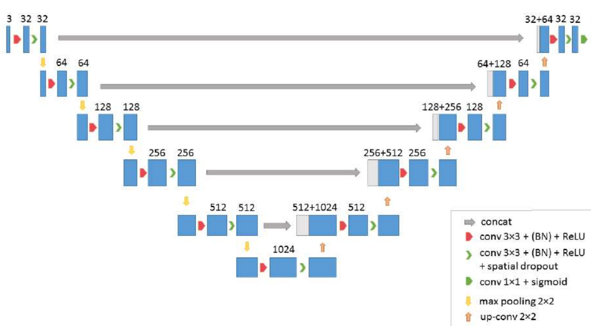

Fig. 4 UNet Architecture [15]

Fully Convolutional Neural Network: We train a FCN to map from an RGB input to a posterior probability map using the architecture. The network contains 35 layers. We fix the stride as 1 and use Rectified Linear Units (ReLUs) as the activation function for each convolutional/deconvolutional layer. Images were resized to 320 x 320 px. For output layer, we use sigmoid function as the activation function. Pixel-wise classification is performed and FCN is essentially served as a filter that projects the entire input image to a map where each element represents the probability that the corresponding input pixel belongs to the lesion. Although this model did not perform well on our dataset.

D. Performance evaluation

All 3 deep learning models were evaluated using following metrics:

Jaccard Index (Intersection Over Union): The Jaccard index, also known as Intersection over Union and the Jaccard similarity coefficient is a statistic used for assessing the similarity and difference between ground truth and model output [10]. The Jaccard coefficient measures similarity between finite sample sets, and is defined as area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth [16], taking α is ground truth and β is obtained segmented mask in output image:

$$\text{Jaccard Index, } j(\alpha,\beta) = \frac{|\alpha\cap\beta|}{|\alpha\cup\beta|} = \frac{|\alpha\cap\beta|}{|\alpha|+|\beta|-|\alpha\cap\beta|}$$

Dice coefficient: It is like IOU range from 0 to 1, where 1 signifies highest similarity between ground truth and predicted output. It is similar to precision as it penalizes for the false positive the methods find, does pixel wise comparison [11]. It is defined as 2 * the Area of Overlap divided by the total number of pixels in both images [16], taking α is ground truth and β is obtained segmented mask in output image:

$$\text{Dice coefficient, DC } (\alpha,\beta) = 2 \times \frac{|\alpha\cap\beta|}{|\alpha+\beta|}$$

Precision: It determines the purity of positive predictions relative to the ground truth, it is useful when cost of false positive is high. It is defined as the number of skin pixels correctly classified (i.e. true positive) by predictions (true positive + False positive) [17].

$$\text{Precision, P} = \frac{TP}{TP+FP}$$

Recall: it refers to sensitivity and is best useful when there is a high cost associated with False Negative. It is defined as the number of skin pixels correctly classified (true positive) by ground truth (true positive + false negative) [17].

$$\text{Recall, R} = \frac{TP}{TP+FN}$$

Accuracy: it simply gives percentage of correctly classified pixel.

$$\text{Accuracy, Acc} = \frac{TP+TN}{TP+TN+FP+FN}$$

E. Key Components Validation

Training the model on 2160 images and validating them with 180 images, the network produced results after running for 150 epochs. Accuracy and Loss measure curves with respect to number of Epochs have been plotted, for training data as well as validation data. We can observe on Figure. 5 that the loss on was around 39% at first then it became 4% as number of epochs increases, and the accuracy was 84% and first then it became 94% on training datasets. On Validation dataset the loss was around 42% which lowered down to 31%, accuracy was 84% which became 94% on increasing number of epochs.

Table 2: Performance Statistics on Test Set

Method	Precision	Recall	IOU	DC	Acc	Loss
SegNet	92.01	87.1	78.13	87.39	94.49	17.29
FCN	89.29	79.23	57.86	72.56	91.89	21.68
UNet	90.87	90.6	81.89	89.66	94.81	28.56

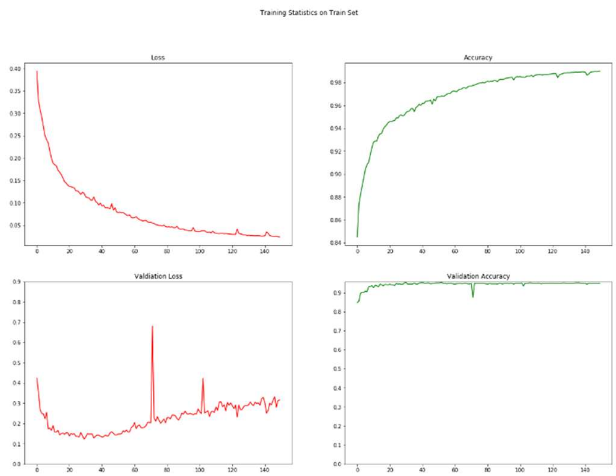

Fig. 5: UNet Training Statistics on Test set

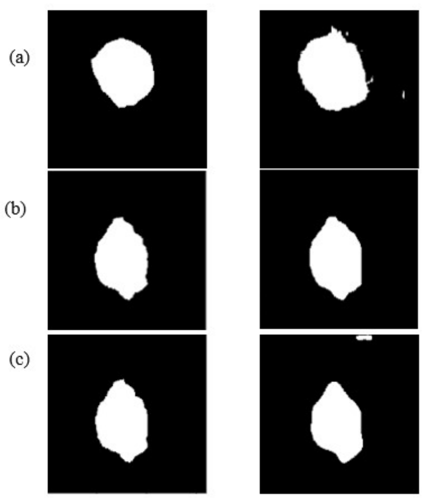

Fig 6: Comparing Predicted Lesion, left: ground truth; right: Predicted (a) FCN (b) UNet (c) SegNet

III. DISCUSSION

The UNet model and SegNet model performed better compared to other state of arts. We can try these models for segmentation of other images as well. We can also augment the data by using other Data Augmentation Techniques like increasing Brightness, Noise Injection, Color Space, etc. but increasing the size space will increase the training time of model. The computation time for our model was 17s 8ms per epoch for training. Post processing was required to enhance the image. Combining different models may produce better predictions without enhancement.

IV. CONCLUSION

For Early detection and diagnosis of melanoma skin cancer, accurate segmentation of skin lesion using automated image analysis is crucial part. Three deep leaning neural architecture were trained and tested on ISIC 2016 Dataset and evaluated. 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. UNet provided the best segmented image mask with Dice Coefficient value of 89.66.

V. ABBREVIATIONS

ID TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative, Acc= Accuracy, DC = Dice Coefficient, IOU = Intersection Over Union, px= Pixel

REFERENCES

1. T. Rath, “Malignant melanoma,” European Surgery – Acta Chirurgica Austriaca, vol. 38, no. 2. Springer, pp. 145–148, Apr-2006, doi: 10.1007/s10353-006-0234-z.
2. A. Kasuya, M. Aoshima, K. Fukuchi, T. Shimauchi, T. Fujiyama, and Y. Tokura, “An intuitive explanation of dermoscopic structures by digitally reconstructed pathological horizontal top-down view images,” Sci. Rep., vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-56522-8.
3. N. Mittal, S. Tanwar, and S. K. Khatri, “Identification enhancement of different skin lesion images by segmentation techniques,” in 2017 6th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2017, 2018, vol. 2018-Janua, pp. 609–614, doi: 10.1109/ICRITO.2017.8342500.
4. F. del Rosario et al., “Performance of a computeraided digital dermoscopic image analyzer for melanoma detection in 1,076 pigmented skin lesion biopsies,” J. Am. Acad. Dermatol., vol. 78, no. 5, pp. 927-934.e6, May 2018, doi: 10.1016/j.jaad.2017.01.049.
5. “ISIC Archive.” [Online]. Available: https://www.isicarchive.com/!/topWithHeader/wideContentTop/main. [Accessed: 02-Jun-2020].
6. R. R. Winkelmann, N. Tucker, R. White, and D. S. Rigel, “Pigmented skin lesion biopsies after computer-aided multispectral digital skin lesion analysis,” J. Am. Osteopath. Assoc., vol. 115, no. 11, pp. 666–669, Nov. 2015, doi: 10.7556/jaoa.2015.137.
7. R. Garnavi, M. Aldeen, M. E. Celebi, G. Varigos, and S. Finch, “Border detection in dermoscopy images using hybrid thresholding on optimized color channels,” Comput. Med. Imaging Graph., vol. 35, no. 2, pp. 105–115, Mar. 2011, doi: 10.1016/j.compmedimag.2010.08.001.
8. I. Pirnog, I. Marcu, and C. Oprea, “Automated segmentation of pigmented skin lesions images for smartphone applications,” in Proceedings of the International Semiconductor Conference, CAS, 2019, vol. 2019-October, pp. 107–110, doi: 10.1109/SMICND.2019.8923938.
9. M. E. Celebi et al., “Border detection in dermoscopy images using statistical region merging,” Ski. Res. Technol., vol. 14, no. 3, pp. 347–353, Aug. 2008, doi: 10.1111/j.1600-0846.2008.00301.x.
10. J. Lameski, A. Jovanov, E. Zdravevski, P. Lameski, and S. Gievska, “Skin lesion segmentation with deep learning,” in EUROCON 2019 - 18th International Conference on Smart Technologies, 2019, doi: 10.1109/EUROCON.2019.8861636.
11. S. S. Kolekar and P. G. Magdum, “Skin Lesion Semantic Segmentation Using Convolutional Encoder Decoder Architecture,” in Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018, 2018, doi: 10.1109/ICCUBEA.2018.8697510.

12. A. Adegun and S. Viriri, “Fully convolutional encoderdecoder architecture (FCEDA) for skin lesions segmentation,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, vol. 11683 LNAI, pp. 426–437, doi: 10.1007/978-3-030-28377-335.
13. P. Brahmbhatt and S. N. Rajan, “Skin Lesion Segmentation using SegNet with Binary Cross-Entropy — Papers With Code,” International Conference on Artificial Intelligence and Speech Technology, 2019. [Online]. Available: https://paperswithcode.com/paper/skin-lesionsegmentation-using-segnet-with. [Accessed: 02-Jun-2020].
14. 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, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.
15. Z. Al Nazi and T. A. Abir, “Automatic Skin Lesion Segmentation and Melanoma Detection: Transfer Learning Approach with U-Net and DCNN-SVM,” 2020, pp. 371–381.
16. “Metrics to Evaluate your Semantic Segmentation Model.” [Online]. Available: https://towardsdatascience.com/metrics-to-evaluate-yoursemantic-segmentation-model-6bcb99639aa2. [Accessed: 02-Jun-2020].
17. F. Gasparini and R. Schettini, “Skin segmentation using multiple thresholding,” in Internet Imaging VII, 2006, vol. 6061, p. 60610F, doi: 10.1117/12.647446.